STATISTICAL AVERAGES

A Methodological Study

BY

DR. FRANZ ŽIŽEK
Privat-Docent in Statistics in the University of Vienna

AUTHORIZED TRANSLATION
WITH ADDITIONAL NOTES AND REFERENCES
BY

WARREN MILTON PERSONS
Professor of Economics in Colorado College

NEW YORK
HENRY HOLT AND COMPANY
1913

THE QUINN & BODEN CO. PRESS
RAHWAY, N. J

TRANSLATOR'S PREFACE

I have long felt the want of a general methodological work to be used as the basis of a college course in statistics. *Die statistischen Mittelwerte,* by Dr. Franz Zizek, seemed to me to meet the requirements of a non-mathematical text-book on statistics better than any work available in English and, consequently, this English translation was undertaken.

Statistical Averages, when used as a text-book, should, of course, be supplemented by lectures, assigned readings, and statistical problems. Lectures and assigned readings should cover such topics as the history of statistics, the organization and work of labor and census bureaus, the preparation of schedules, the accuracy obtainable in statistics, the construction and use of diagrams and maps, and the mathematical methods of computing the standard deviation and the coefficient of correlation. Numerous problems should be assigned so that the students may become familiar with frequency tables, graphic representation, the methods of computing the important index numbers (such as the Labor Bureau index numbers of wages and prices and the *Economist* index number of prices), the standard deviation, correlation tables, the coefficient of correlation, and the like. By the assignment of problems the students will become familiar with the sources of statistical data and with the facts of statistics as well as with the methods of statistics.

Perhaps the most important advantage that *Statistical Averages* offers to American readers is to be found in the explanations of, and copious references to, statistical data and methods from French, German, and Italian

sources. Probably the chief defect of the book is the lack of illustrative matter in the way of graphic representation and statistical tables of various sorts. However, this defect can be turned to advantage in class-room use by requiring the students to secure such matter and present it, graphically and otherwise, as indicated in the preceding paragraph.

A few additions have been made to the translation in the way of footnotes signed "Translator," new titles added to the bibliography, and selected references to those statistical journals and publications which are of especial use to American readers.

In conclusion, I acknowledge my obligation to Dr. Franz Zizek, who cheerfully authorized the translation and who corrected the manuscript, to Mr. J. C. Schwartz, who aided in the translation, to Professor T. S. Adams of the University of Wisconsin, who read the proof, and especially to Professor W. K. Stewart of the German Department of Dartmouth College, who prepared a considerable portion of the translation and who aided in other ways.

W. M. P.

CONTENTS

PART II

THE VARIOUS KINDS OF AVERAGES

PART III

DISPERSION ABOUT THE MEAN OR AVERAGE

APPENDIX I

APPENDIX II

APPENDIX III

STATISTICAL AVERAGES

INTRODUCTION

In nearly all branches of statistical investigation averages or means (French, " moyennes," German, " Mittelwerte," Italian, " medie ") are very often used for various purposes of the greatest significance. Averages are, therefore, undoubtedly to be reckoned among the most important aids in statistical method. Even Guerry, the founder of moral statistics, indicated the extraordinary importance of averages in the following definition of statistics: " The science of statistics consists essentially in the methodological enumeration of variable elements, whose mean it determines." Edgeworth defines statistics as " the science of those means which are presented by social phenomena "[1]; and similarly Bowley says, " Statistics may rightly be called the science of averages."[2] The application of averages has, as is well known, given rise to controversies of various kinds, which fill a considerable part of statistical literature. Not infrequently the incorrect use of averages has also led to erroneous conclusions and to contradictions, which have shaken our confidence in statistics. Averages, indeed, are only applicable under strictly defined conditions, and conclusions based on averages are likewise permissible only within well defined limits. It is the task of statistical science to investigate the application and use of averages from the general methodological standpoint, and to determine the part which averages should play in statistical method.

[1] " On Methods of Statistics," Jubilee Volume of the Royal Statistical Society (1885), p. 182.

[2] Elements of Statistics, 2nd ed. (1902), p. 7.

Statistical literature does, indeed, possess a large number of works which deal with isolated questions connected with the application of averages or with the various averages in use in definite departments of applied statistics (for example, the average length of life, or average number of children per family). Especially the " mathematical statisticians " (Lexis, Edgeworth, Westergaard, von Bortkiewicz, Pearson, Galton, Yule, Bowley, and others), as well as some philosophers and theoretical mathematicians, who have also occupied themselves with statistical problems (as, for example, Fechner, J. von Kries, Czuber, and Blaschke), have thoroughly investigated various methodological questions connected with averages by using the calculus of probabilities.[3] But only a few works have for their object the treatment of statistical averages in the most general methodological manner.[4] As a matter of fact, not one of the treatises referred to offers anything like an exhaustive development of the problem. Moreover, most of these works are decidedly out of date.

Furthermore, the discussions regarding averages, which are to be found in the numerous handbooks, text-books, and

[3] The mathematical investigations of formal theories of population and the measurement of mortality by such authors as Becker, G. F. Knapp, Zeuner, and Wittstein have little bearing on our problem.

[4] Among such independent studies are:

Bertillon, Adolphe, "La théorie des moyennes en Statistique," Journal de la Société de Statistique de Paris, 1876.

Bertillon, Adolphe, article entitled "Moyenne" in the Dictionnaire encyclopédique des sciences médicales.

Edgeworth, F. Y., article entitled "Average" in Palgrave's Dictionary of Political Economy.

Fechner, G. Th., Kollektivmasslehre, published by G. F. Lipps, Leipzig, 1897.

Holmes, George K., "A Plea for the Average," Quar. Pubs. of the Am. Stat. Assoc., New Series No. 16, December, 1891.

Messedaglia, Angelo, "Il calcolo dei valori medî e le sue appli-

systematic treatises, as well as in the investigations devoted to statistical method in general, are for the most part rather meager. General works on statistics and statistical method cannot, in the nature of things, give much space to the individual problems in which statistics abound.[5]

In the following pages the attempt will be made to offer a systematic treatment of the most important questions which make up the problem of statistical averages. The scope of these questions is extremely wide, as a very great number of statistical methods depend upon the application of averages, and as the most important objects of statistical investigation demand the use of averages. For these reasons the problem of averages may be correctly designated as one of the most important in scientific statistics, and in a sense as the central problem of statistics. Since statistics by its nature deals with phenomena, both variable and complicated, it is evident that averages, which characterize such phenomena by a single number, must be of preeminent importance in the science.

Our aim in the following treatment is a general methodological one. That is to say, our aim is to determine those properties which the various types of averages, such as, arithmetic mean, geometric mean, median, mode, etc., possess intrinsically, irrespective of the department of statistics (population statistics, economic, moral, biological sta-

cazioni statistiche," Arch. di Stat., Anno V, 1880. The same in French: "Calcul des valeurs moyennes," Annales de démographie internationale, IV, 1880.

Quetelet, A., "Sur l'appréciation des moyennes," Bull. Commiss. Central. Statist., Vol. II, 1845.

Quetelet, A., Lettres sur la théorie des probabilités, Pt. II, "Des moyennes et des limites," 1845.

Tammeo, G., Le medie e loro limiti, 1878.

Venn, J., "On the Nature and Use of Averages," Jour. of the Roy. Stat. Soc., 1891.

[5] See Appendix III of this book for a list of titles of general works on statistics.

tistics, etc.) to which the numerical data, from which the average is computed, belong. We shall show that the same methodological problems constantly recur in the most diverse fields of statistics, and that these problems have a common solution, although heretofore each field has been worked independently without aid from the results obtained in other fields. It is also important to note that this method of handling averages is in accordance with the most significant evolutionary tendency of modern statistics.

The author does not possess sufficient mathematical training to be able to apply or examine critically the methods of "mathematical statistics," which—apart from the theory of the development and growth of population—are all connected with the problem of averages. Nevertheless, such methods will not be entirely disregarded. On the contrary, in order to supplement the exposition of the elementary mathematical methods, it will be shown upon what fundamental principles tne methods of "mathematical statistics" depend, to what problems these methods have been applied. and what interesting results have been attained. The author deems some consideration of "mathematical statistics" indispensable because its problems do not differ essentially from those of elementary scientific statistics. By the application of the calculus of probability the "mathematical statistician" attempts, with mathematical precision, to solve problems which confront any scientifically minded investigator. Since the majority of statisticians are trained in economics and not in mathematics the author will attempt to give in non-technical language some information of the processes and results of mathematical investigations in statistics. In this way the significance of the calculus of probability in general statistical method will become apparent.

PART I

STATISTICAL AVERAGES IN GENERAL

CHAPTER I

CLASSIFICATION OF STATISTICAL SERIES WITH REFERENCE TO THE PROBLEM OF AVERAGES

Statistical series are classified in various ways in the textbooks of statistics. They are usually differentiated as space, time, and qualitative or quantitative series, according as the individual members of the series are distinguished by their distribution in space (geographical divisions) or time, or by their qualitative or quantitative differences. It is also customary to classify statistical series according as their individual members are absolute numbers, or relative numbers or averages. With reference to the problem of averages a classification of a particular kind is appropriate. The various series are, therefore, embraced in three groups according to the different methods of ascertaining the averages. These three groups must also be differentiated in the discussion of the different special problems.

In the *first group* are to be included those series of observations upon individuals or units of various kinds which, for the purpose in mind, are treated as similar. In these series each individual member refers to an observation unit which is marked by some defined character. This character may be qualitative or quantitative. We have to deal with a qualitative character when, for example, the sex or the occupation of certain individuals is being considered.

Since, however, qualitative individual observations do not permit the computation of an average they do not enter further into our problem. For that reason we shall deal exclusively in the following pages with quantitative in-

dividual observations. Such observations arise ordinarily through measurement. Thus, for example, the age, wages, income, length of life, etc., of single individuals in definite groups of the population are measured and are then represented in the form of series. Cases occur, however, in which the items contained in the series do not deal with real measurements but with quantitative observations of another kind, such as are obtained by counting. Thus, for instance, houses are observed with reference to the number of their occupants, families with reference to the number of children.[6]

From series of quantitative observations various kinds of averages may be computed, of which the most important are the arithmetic mean, the median (that is, the middle number of the series when the items are arranged accord-

[6] Series of quantitative individual observations, whether they be actual measurements or quantitative individual observations of another kind, are either space, time, qualitative, or quantitative series. The units to which the observations refer frequently belong to different space or time divisions; thus, for example, the domiciles of persons whose age, income, etc., are measured present space differences, and the data giving ages at marriage or death present time differences. But the space or time differentiation of the individual observations, which appears in the original material, normally disappears during the course of the statistical work and is not evident in the resulting statistical series. Series of quantitative individual observations are, furthermore, not qualitative series, since similar units are selected for measurement. Neither are series of quantitative individual observations identical with the group of quantitative series—as many authors appear to assume—because only those series are to be designated as "quantitative" whose items are differentiated from one another by some quantitative criterion, as, for example, is the case with series of death rates, birth rates, etc., for different age classes of the population. The fact that quantitative individual observations possess different numerical values for the element of observation does not constitute them quantitative series, since all series (time, space, etc.) consist of numbers of various sizes. The above-mentioned customary division of statistical series into space, time, qualitative, and quantitative groups is, therefore, not exhaustive.

ing to size, or, in case there is an even number of items, the arithmetic mean of the two middle numbers), and the mode (that is, the relatively most frequent value, the point of greatest density). The average computed from the series represents the mean of the measurement in question (average age, average wage, average income, mean and probable lifetime); or else it indicates, when a definite quantitative character is obtained by counting, how many units of that character occur on the average (for instance, the average number of occupants per house, or children per family).

Quantitative observations are not always presented with the greatest possible detail. Often the variant items are tabulated according to class (for example, age, income, wage classes, etc.). The frequency table, thus obtained, merely indicates (absolutely or relatively) how many items belong to the different classes. Averages may be computed from the frequency tables for the character in question, either measured or counted, no matter whether the tables consist of absolute or relative numbers.

The items which produce the series in question may belong to the most varied branches of social life. Moreover, similar series may arise from observations in natural science. Especially meteorological (thermal and barometric) observations, and also anthropological measurements (height, chest-girth, various dimensions of the skull, muscular power, lung capacity, etc.), produce series which are well adapted to the application of statistical methods. Likewise, series of measurements of certain characters of animals and plants have recently been investigated according to the methods of scientific statistics. Indeed, the most important methods of modern mathematical statistics have been developed from biological material, and statistical method plays as important a part in modern biology as it does in sociology. In particular, the questions of variation and heredity are being investigated with great success

by use of the statistical method. Instead of speaking of individuals we may, therefore, speak, with Gustav Theodor Fechner, of "collective objects."[7] Fechner understands by a collective object one which consists of an indefinite number of accidentally varying specimens, which are grouped by a generic notion. Man, in general, according to Fechner, forms a collective object in the wider sense; man of a definite sex or age forms a collective object in the narrower sense. Meteorological observations, anthropometric data, measurements of animals and plans, etc., represent other collective objects whose chance variation the science of collective masses ("Kollektivmasslehre") investigates.

All the illustrations quoted have had to do with series of observations which refer to different individuals with some common characteristic. The items of a series, however, may arise from repeated observations, especially measurements, of the same object. The mean computed from such series is called "objective," in contradistinction to the "subjective" mean computed from series of single observations of a number of units.[8]

Statistics, as a social science, deals constantly with series of single observations of various similar units and, therefore, with "subjective" means in the above sense. We take the wages, the income, the age of *various* individuals and compute the average. Likewise, the means computed from meteorological observations, anthropometric data, and

[7] Cf. Kollektivmasslehre by G. T. Fechner. Lipps and Bruns may also be mentioned as supporters of "Kollektivmasslehre."

[8] Cf. A. Bertillon, "La théorie des moyennes en Statistique" in the Journal de la Société de Statistique de Paris, 17th year, p. 266; also J. Bertillon, Cours élémentaire de Statistique administrative, p. 112, and G. v. Mayr, Theoretische Statistik, p. 98. Block (Traité théorique et pratique de Statistique, 2nd ed., p. 129) rejects the classification of averages into "objective" and "subjective" means for the unsatisfactory reason that it is necessary to reserve the notion of a "moyenne typique" for the means "qu'on prend sur

biological measurements are " subjective." On the other hand, in other fields, especially astronomy and geodesy, repeated measurements are often made of the same object, in which cases an " objective " mean is computed. For example, the problem may be to determine the declination of a star or the zenith-distance of its path across a definite meridian, or else to measure the latitude of a place or the distance between two points. By computing the arithmetic mean of repeated measurements the " most probable " size of the object in question is obtained. This average, in all probability, gives most nearly the true, or ideal, size of the object whose measurements are affected by accidental errors.

It was with reference to such " objective " means that the principles of the theory of errors were developed, especially by Gauss. These were first applied by Quetelet to series of anthropometric data, and later, by mathematical statisticians, to other series of individual observations. Such studies have shown that the characteristic distribution of the items about their mean, which marks the series of measurements of the same object (called " normal " distribution), holds sometimes also for series of measurements of similar units. Where this is the case, the similar items may be regarded as empirical values, affected by accidental errors, and in such cases the average, which represents the ideal value, gains additional significance. This average may be regarded as the resultant of the

une série de mesures opérées sur un même objet, ou qui ne s'appliquent qu'à des grandeurs peu différentes." Block believes, then, that means of series of repeated measurements of the same object and means of series of small dispersion may be classed together as "typical" means. But the distinction between series according as they originate from repeated measurements of the same object or single measurements of different objects should not be confused with the distinction between series according to the kind of their dispersion. The notion of a "typical" mean should be reserved for means of series with a definite dispersion.

complex of general causes producing the single items, and is a "typical" mean in the strictest sense of the word.

Thus the mathematical theory developed from repeated measurements of the same object affords a special basis for the arithmetic mean of certain series of single measurements of various similar units and gives us a standard for judging the dispersions of such series. It is to be noted, however, that the Gaussian law of normal distribution holds for series of measurements of different units only in isolated cases. Anthropometric series of this kind have sometimes been established, especially measurements of height. Lexis has proved that the dispersion of items about the "normal" length of life is in accordance with the theory of errors. On the other hand, Fechner and, especially, Pearson have noted numerous statistical series which do not correspond to the Gaussian law, but which, in spite of their asymmetry, may be brought under a generalized law of accidental variation. In general, then, it appears that the normal distribution holds for repeated observations of the *same* object, but that the series of observations of *different* objects, as a rule, show an unsymmetrical distribution about the average.

A further distinction between series of measurements of the same object and of similar objects may be noted. In series of repeated measurements of the same object we are concerned with establishing the true size of the object with the greatest possible accuracy. Such series follow, as has been said, the Gaussian law, and the three means, the arithmetic mean, the median, and mode, theoretically coincide. In series whose members refer to different, even though similar, objects, it is not proper to speak of a "true" value. All measurements of such objects, however much they may differ from their mean, are equally true. It is, therefore, simply a question of applying the most suitable and comprehensive descriptive term to the group of meas-

urements. For this purpose the statement of the average or averages is indispensable. If the series follows the Gaussian law, then, even in the case of single measurements of different units the three means, above mentioned, coincide. If, on the contrary, the series does not follow this law then the three means diverge from one another and should, if possible, all be stated, since each of them contributes something unique towards characterizing the series.

In consequence of this essential distinction between objective and subjective means, a different importance is to be attached to the dispersion about the means of the two series. The dispersion of a series of measurements of the same object depends exclusively upon the accuracy of the measuring instruments and determines the degree of precision of the mean. The magnitude that, added to the arithmetic mean, and subtracted from such mean, gives two limits between which a *given* proportion of the measurements fall, may be taken as an index varying inversely with precision of the instrument.[9] The dispersion of series of measurements of various units plays a more substantial rôle, since each item of such a series is to be regarded as a fact determined by peculiar causes. The dispersion of the series, therefore, indicates the variability of the phenomenon in question.

Consequently, the distinction between subjective and objective means is of importance in several directions. If we conceive statistics merely as a social science, then it follows that the "objective" mean plays no part in it. If we, however, investigate the theoretical foundation of the statistical method, then it is necessary to become familiar with the mathematical principles which have developed from consideration of repeated observations of the same object, and which are now also applied to the "subjective" mean. In the latter case we must, of course, investigate the modifications required in applying these principles to

[9] Cf. G. Th. Fechner, Kollektivmasslehre, p. 15.

series of measurements of similar objects taken from the fields of sociology, biology, meteorology, etc.

A *second group* which we wish to differentiate may be illustrated by the population figures for the various districts of a country. The members of such series give the size of definitely limited groups, aggregates, or masses (e. g., counties), which, taken together, form a totality of a higher order (e. g., state). Stated in general terms, the second group embraces those series whose members give the size of definitely limited masses [9a] which, taken together, form a totality of a higher order. The members of such series are not, like those of the first group, similar items, either measurements or other elements of observation, but they are statistical masses, mutually limited as to size by the point of view taken in the particular problem. The definition of the masses may be from the point of view of space, time, qualitative, or quantitative differences. The average computed from the series gives the mean value of the masses belonging to it, that is, the average number of units per mass.

If, for instance, we posses the population figures for the various districts of a country (space masses), we may then determine the average population per district. If we have the number of births or deaths for a series of years (time masses), we may reckon how many births or deaths occur

[9a] The word "mass" has been used throughout this translation to designate a group of units. The science of statistics, of course, passes from consideration of groups of concrete units to detailed examination and analyses of abstract figures characterizing such groups. Thus, the science deals directly with series of items. However, the results obtained from examination of abstract figures or items are made useful only by connecting them with some concrete mass. In this way statistics becomes an applied science of masses. The use of the word "mass" in this translation accords with its use in the name which has been given to the mathematical instrument which has been invented for rendering statistical data available for scientific purposes, i. e., "the calculus of mass phenomena" (see, for instance, H. L. Moore's Laws of Wages, p. 4).—TRANSLATOR.

on an average per year of the given period. If we have data as to the sizes of the different occupations (qualitative masses) or of age classes of the population (quantitative masses), we may find out the average number of persons per occupation or age class. But such averages computed from qualitative and quantitative series are, as a rule, worthless, since the average size of a mass depends solely on the number of masses, the statistician himself generally determining arbitrarily the number of qualitative or quantitative masses into which he divides the totality of the higher order. For example, the greater the number of occupation or age classes, the fewer will be the number of people assignable to a single one.[10] On the other hand, when the statistician deals with space or time, he is handling objective magnitudes (years, districts, etc.), and the computation of the average size of a mass in such a sense may be of great significance.

From the series of the second group, as a rule, only the arithmetic mean is taken. The prerequisites for the application of the other means do not normally occur.[11]

[10] For similar reasons it is useless to compute the average size of a constituent from a series of percentages which gives the space or time sizes of the constituents in relation to a totality of a higher order. The size of the average of such a series depends solely upon the number of constituents. For example, suppose we have statistical data for ten years or classified for ten sections of a country, or if we divide the population into ten occupation groups or into ten age classes, then one year, one section, one occupation group or one age class would, of course, contain, on the average, 10% of the cases.

[11] The quantitative series of the second group are, considered from another point of view, at the same time members of the first group, that is, they are series of individual observations which are embraced by the quantitative constituents (magnitude classes). A series, for example, which gives the distribution of the population according to age class consists of individual age data combined into magnitude classes. The same holds true for series showing the number of persons receiving various classes of income or wages. If we consider such series as series of the second group and com-

A *third group* consists of those series whose members are not absolute but relative numbers. These relative numbers may be either subordinate or coordinate. Subordinate numbers are those which indicate the relative size of parts to a whole, usually by percentages; for example, the percentage of male and of female births to the total number of births, or the percentage-distribution of deaths according to sex, age, or social class. Coordinate numbers are those which indicate the relative size of pairs of coordinate masses. They originate through the interrelation of two such masses, namely, by dividing one by the other and sometimes multiplying the quotient by 100 or 1,000. Thus the coordinate number, obtained by multiplying the ratio of the number of deaths to the number of living by 1,000, indicates the death rate per thousand. In the same way, we may obtain the marriage rate or birth rate per thousand of population. Neither subordinate nor coordinate numbers arise from simple measurement or counting, but always from computation.[11a]

The masses, which are characterized by the relative numbers, may have originated from a criterion of space, time, quality, or quantity. Thus we may, for instance, represent the sex distribution of infants by a series of subordinate numbers for various districts, months, religious denominations, or age classes of the parents according as we use a criterion of space, time, quality, or quantity in dividing

pute the mean size of a quantitative constituent, as explained above, we obtain averages of less significance than if we consider them to be series of the first group and compute the mean size of the element of observation (age, income, wage, etc.), which is of the highest significance.

[11a] When relative numbers are stated in a numerical form such that if used to multiply a total (e. g., population) an allied number (e. g., number of births) will result, then such relative numbers are called "statistical coefficients." "Thus, if the birth-rate is 40 per 1,000, the coefficient is .04" (Bowley, Elements of Statistics, p. 129).—TRANSLATOR.

the total number of births in a year, and compute the sex-ratio for each of these divisions. Similarly we may secure death rates for various districts, months of the year, religious denominations, or age classes by taking the ratios of the number of deaths to the correlated population, both deaths and population being differentiated according to a criterion of space, time, quality, or quantity. These ratios form a series of coordinate numbers.

It is characteristic of the third group of series that the items generally refer to masses of varying size and therefore of varying significance or weight. But the relative importance is not apparent from the items themselves. It is of the nature of relative numbers that they do not express the absolute size of the masses from which they originate; 1,000 men and 1,000 women give the same sex ratio as 1,000,000 of each, and 50 deaths among 2,000 living give the same ratio as 2,500 deaths among 100,000 living.

Whether the series be composed of subordinate or of coordinate numbers, whether the criterion be that of space, time, quality, or quantity, it is evident in general that items arising from different districts, different years, different occupations or different age classes must produce relative numbers of varying weights. Because of the varying weights, a mean should not usually be computed directly from items of the third group, as it was in the cases of the first and second groups. On the contrary, the mean of such series is independent of the individual items, and it must be computed from the original data which gave rise to the items of the series. Thus, in order to get the true average annual death rate for a period of 10 years, the total number of deaths during such period must be divided by the sum of those living at the beginning (or other definite point of time) of each of the 10 years.

The mean is therefore not to be computed from the series

but, found independently, it supplements the items of the series. The original data are used, then, in whole or in part, to compute the individual items of a series and their mean, both items and mean being relative numbers. In point of time the computation of the items may sometimes follow the computation of the average value for the totality. In fact the tendency of statistical research has been as a first step to ascertain general averages and secondly to separate the statistical data into more homogeneous parts, for which separate relative numbers are computed. Thus, general death rates are followed by death rates for various age classes, occupations, etc.

Since the mean and the items are computed independently, it may also happen that not all of the components of the total are taken. For example, the probability of death may be computed for the whole population and for certain easily defined occupations, while other occupations are disregarded.

From what has been said, it follows immediately that the computation of the different means, possible for the first group, is not possible for the third. The general average originates from the absolute numbers found for the totality. These absolute numbers are the sum of the component absolute numbers, from which the items were computed. The general average is thus really obtained by weighting the individual items, and consequently it may be computed directly from such items by finding the weighted arithmetic mean. Consequently, if the original data are not available, we may compute a weighted arithmetic average from the items of the series, assigning weights as nearly correct as possible. If we only know that the respective weights are not essentially different (such as may occur in a time series pertaining to a population which does not change essentially during the period considered), we may be contented with a simple arithmetic average of the items of the series. In

case the weights are identical the simple arithmetic mean of the items will coincide with the general average computed independently from the original data.

In anthropometry several relative numbers are used (for example, the ratio of the length of the head to its breadth, or the cephalic index) which differ from the relative numbers computed from demographic data in that they originate through correlating, not masses, but single measurements. Averages are frequently computed from such relative numbers; for example, average cephalic index. As the dividends and divisors, which give the ratios, differ among themselves the average computed directly from the various ratios is different from the quotient of the sum of the dividends and of the divisors. An illustration may be taken from Dr. Bertillon's "*Théorie des moyennes.*"[12] If we measure two skulls—limiting ourselves to two for sake of simplicity—and find the first 200 mm. long by 180 mm. broad, and the second 160 mm. long by 112 mm. broad, the cephalic indices will be 90 and 70, respectively; giving an average index directly computed of 80. But dividing the sum of 200 and 160 by 180 plus 112 we obtain the average 81.2. Mathematicians and anthropologists differ as to which method is more correct. The results obtained by the two methods differ, as a rule, but little.[13]

Those relative numbers have a particular significance which are expressed as numerical probabilities, or known functions of them. Lexis defines a statistical probability as a fraction whose numerator gives a number of observed special cases or elements, which either *originate from* the number of observed cases or elements given in the denominator, or are actually

[12] Journ. de la Soc. de Stat. de Paris, 1876, p. 314.

[13] Cf. Chap. XXII of Fechner's Kollektivmasslehre: "Kollektive Behandlung von Verhältnissen zwischen Dimensionen; mittlere Verhältnisse" (§§147-151, pp. 352-364).

included in such denominator.[14] He calls the former case of probability relation genetic or primary, the latter, analytic or secondary. "In the former case the numerator gives the number of events of a particular kind which originate from the totality forming the denominator; for instance, the ratio of death in a definite age class to the number living who have reached the lower limit of the class in question and have been subject to the death risk in question."[15] "In the case of analytic probability relation, on the contrary, the units of the numerator are of the same kind as those of the denominator and are only distinguished by some special characteristic; the numerator, therefore, forms a section of the totality expressed by the denominator. Such is the ratio of the number of male births to the total number of births."[16] Genetic numerical probabilities (for instance, the probability of death) are, in our terminology, coordinate numbers; analytic numerical probabilities correspond to subordinate numbers (for example, the probability of a male birth corresponds to the percentage that the male births bear to the total number of births).

As has been stated, functions of probabilities must also be considered. An example of such is the ratio of male to female births, that is, the coordinate number indicating how many male births occur to 1,000 female births. This coordinate number is in itself not a probability but it is a function of such, that is, a function of the (analytic) probability of a male birth in relation to the total number of births. If we let $v =$ the probability of a male birth, $p =$ the ratio

[14] Abhandlungen zur Theorie der Bevölkerungs- und Moralstatistik, IV, "Übersicht der demographischen Elemente und ihrer Beziehungen zu Einander," p. 62. Cf. also von Bortkiewicz, Das Gesetz der kleinen Zahlen, p. 26, and Czuber, Wahrscheinlichkeitsrechnung, p. 302 f.

[15] Such a probability applies directly to the success or failure of an event.

[16] Cf. Abhandlungen zur Theorie, etc., V, "Über die Ursachen der geringen Veränderlichkeit statistischer Verhältniszahlen," p. 84.

of the male to the female births, $z = 1,000$, $p =$ the number of boys born to 1,000 girls, then the equation $z = \frac{1000v}{1-v}$ follows. If on the average 515 males occur in every 1,000 births, then the probability of a male birth $(v) = 0.515$ and $z = 1,062$, that is, 1,062 males are born for every 1,000 females.

Relative numbers, which do not take the form of numerical probabilities and are not functions of such, are designated by Lexis as "Koordinations-verhältnisse" (coordinate relations). "They are, in general, relations between statistical totals, which are independent from each other either in whole or in part. To this class belong the various coefficients (death or marriage coefficients), and relations such as the yearly number of births to marriages, etc."[17]

The peculiar feature of relative numbers, which appear as numerical probabilities or functions of such, consists in the fact that the methods of the theory of probabilities may be applied to them, while such methods may not be applied to other relative numbers. The theory of probability offers first of all a means of determining the reliability of numerical probabilities, that is, of establishing between what limits the theoretical probability, which gives rise to the observed values, probably lies. Since the reliability of an observed probability varies directly with the number of observations upon which it is based, probabilities based upon larger statistical masses have more scientific value than those based upon smaller ones. Numerical probability determined for a totality is, however, the arithmetic mean (simple or weighted) of corresponding values for its portions. Accordingly, the theory of probability under definite conditions affords a particular *raison d'être* for the arithmetic mean of certain items. Furthermore, it is possible by means of the theory of probability to compare several numerical probabilities as to whether the difference between them may probably be ascribed to chance or whether we must assume

[17] Ibid p. 84 f.

that unequal theoretical probabilities are involved in the two observations; this latter case would indicate essential differences in the causes of the phenomena. Finally, the theory of probability gives a criterion for measuring the dispersion of a series of numerical probabilities, especially for determining whether the various members of the series may in fact be regarded as empirical determinations of the same theoretical probability, or, as the case may be, of a theoretical probability affected only by accidental changes.

More recent authors like Lexis and Bortkiewicz, it is true, go so far as to assert that the various relative numbers, even if they satisfy the formal conditions in question, may not be summarily treated as approximate values of probabilities. They claim that the question, whether definite relative numbers may be regarded as empirical numerical probabilities, can only be answered in the affirmative, if these relative numbers belong to a series of values whose dispersion follows the theory of probability.

Very similar to the series of relative numbers are the series whose members are averages for an individual measurement or other quantitative observation. Such series are not frequent, but do occasionally occur. A large mass of quantitative observations may be subdivided following a criterion of space, time, quantity or quality, and accordingly particular averages may be computed for the parts and represented in a series. Thus, for example, the age data at time of marriage or death may be subdivided and particular values computed for the average age of those marrying or dying in different months or provinces or occupations. The question now arises, how from such series, whose items are themselves averages, we may compute the general average which is necessary for different purposes, especially as a basis of comparison.

The parts, to which the averages of the series refer, are not, as a rule, of equal size. Thus different provinces and different occupations generally show varying numbers of

deaths and marriages, besides, these numbers vary from month to month. Therefore various significance or weight attaches to the items of the series. But averages for an individual observation have, like relative numbers, the characteristic of not expressing the magnitude of the masses to which they refer. From the figure for an average age it is not possible to deduce the number of persons whose age was considered in the computation of the mean. Hence, from a series of averages it is impossible, as a rule, to compute the general average directly; on the contrary, we must compute the more comprehensive average of the higher order independently on the basis of all the individual cases entering into the various subdivisions making up the series.

The relations existing between the means for the totality and for the subdivisions depend upon the kind of averages composing the series. The general average of a series of arithmetic means represents the weighted average of such items. If the original individual data are not available, the average of the higher order may be computed directly from the items of the series, by combining them with properly chosen weights. It may be necessary to estimate the weights. If the items of the series refer to subdivisions of equal weight, the weighted average coincides with the simple arithmetic mean of the items and, therefore, we may dispense with weights.

Of course not only arithmetic means but also other averages such as medians or modes may occur in the series. It may be instructive to compare the items of such series with a general average (median, mode, etc.) computed from the original individual data. Thus we may compare the medial ("probable") or modal ("normal") length of life of different sections of the population with a similar average computed for the total population. However, there is no general rule for the relation existing between the median or mode of a whole and the medians or modes,

respectively, of the parts.[18] This relation will, in every case, depend on the manner in which the whole has been subdivided.

[18] The essentials of the classification of statistical series, given above, rests upon that given by Edgeworth and Czuber. They have not worked out the details but their work suggests them. Edgeworth says: "Two cases may be distinguished: 1, where the returns with which we have to deal are *measurements* in *space* or *time*, e. g, statures of men or ages at death, and *mere numbers*, e. g., yearly deaths; or 2, *ratios;* such as that of male to female births, or rates of mortality" ("On Methods of Statistics," Jubilee Volume of the Roy. Stat. Soc., 1885, p. 188). Czuber says: "As aids in the description of human phenomena there are, besides the relative numbers which we previously considered and which we designated as *intensive magnitudes* because they express the intensity of an action or the frequency of a phenomenon, also *extensive magnitudes* which are expressed in familiar units (dollars, yards, years). In particular, they express the space of time during which a condition exists or they measure the interval between two changes of condition and are partly of biological, partly of sociological and economic interest. As illustrations we may cite: the length of life of individuals of a deceased population; the present length of life of individuals of a living population; the ages of brides and grooms at marriage; the duration of married life; the duration of the activity of individuals in certain occupations; the duration of sickness in various age groups, etc." (Die Wahrscheinlichkeitsrechnung und ihrer Anwendung auf Fehlerausgleichung, Statistik und Lebensversicherung, 1903, p. 333).

CHAPTER II

ISOLATED AVERAGES AND STATISTICAL EVOLUTION FROM ISOLATED AVERAGES TO AVERAGES BASED UPON SERIES OF ITEMS

An average serves to characterize a number of divergent quantities by a single value. These quantities usually form a series. There are, however, also statistical values of the character of averages, which do not originate from series, since the quantities, which the average properly represents, are not known. These " isolated " averages are computed or estimated.

A. ISOLATED AVERAGES OBTAINED BY COMPUTATION

The arithmetic average of a series is the ratio of the sum of the items to their number. In case we do not know the single items, but merely their sum and number, the ratio gives an " isolated " average. Thus, for instance, the average wage of workmen in an industry may be computed, even if the individual wages are not known, in case the total wages and the number of workmen are known. Of course a complete series of wage data would give vastly more information concerning the wage status of the workmen than would an isolated average. The series would show the wage conditions in detail; wage classes could be formed and, besides the arithmetic mean, other means (median, mode, etc.) might be determined. Also the wage data might be tabulated according to sex, age, occupation, *ad lib.* Isolated averages for magnitudes which permit of individual measurement belong, therefore, to a rather primitive stage of statistical method. Especially in the

field of wage statistics the development has generally been from the computation of isolated averages to the complete tabulation of the wages and computation of averages from the individual observations.[19-19a]

[19] For instance, an isolated average wage was computed by dividing the total wages by the number of workmen in Czornig's "Industriestatistik der österreichischen Monarchie für das Jahr 1856." Likewise, the average income per year is obtained in the current Austrian mine-workers' wage statistics (published annually) by dividing the total wages paid by the average number of workmen. This average yearly amount received, however, is differentiated according to the district and occupation of the workmen.

Average wages are also obtained in the United States census by the summary methods under discussion. The total amount of wages paid and the average number of workmen are asked for and the average wage then computed. The inadequacy of this treatment of wages has often been remarked by the Census Office. The objections to the method are given at length in Part I of the Report on Manufactures of the Twelfth Census (Vol. VII, pp. cxi and cxii). Consequently, a special wage investigation was undertaken in connection with the census of 1900, in which detailed wage data were obtained for a large number of industries (see Twelfth Census, Special Reports, Employees and Wages, 1903).

Numerous statisticians, Boehmert in particular, have designated the purpose of wage statistics to be expressly that of dealing adequately with individual earnings. Likewise, the International Statistical Institute expressed a similar opinion in a resolution adopted in 1891 at its Vienna session. But the treatment of rates of wages (taux de salaires, Lohnsätze) also possesses great economic significance The number of workmen receiving certain rates of wages may be ascertained and the average wage rates for greater groups of workers may be computed. From the wage rates and length of time worked the earnings may be derived. The English and American statistics are frequently based upon rates rather than upon earnings. The above-mentioned Report on Employees and Wages was chiefly based upon rates of wages. Similarly, the wage statistics of the Berlin Statistical Bureau, published in 1904, had to do with wage rates for the various kinds of work in various industries.

[19a] Wage groups were given in the reports of the census of 1890. The reports of the bureaus of labor of Massachusetts, New Jersey, and Kansas also present wage groups. Bulletin 93 of the Bureau of

In other cases a similar development is impossible and in these the isolated average is still computed. The reason for such lack of development is either that it is impossible to obtain the individual measurements or that to do so would be inquisitorial on the part of the state. Thus we generally compute per capita consumption of meat, beer, tobacco, etc., as isolated averages.[20] However, if we could obtain the complete series representing the individual consumption of meat, beer, tobacco, etc., new and important facts of value to the economist and hygienist would undoubtedly be ascertained. All those individuals who do not consume these articles and who strongly influence the general average could be excluded; the average per capita consumption could be found for the remainder of the population, which could also be subdivided according to the degrees of consumption. It is not profitable to press these points, however, as these statistics cannot be ascertained.[21] In consumption statistics we must, therefore, be satisfied with isolated averages. Similar averages, likewise, are to be found in other fields. Thus we compute the number of letters, periodicals, or money orders per head, or lottery stakes per head. Although the individual statistics might be of interest it is impossible to obtain them.

Besides the averages which represent a series of measure-

the Census gives the earnings of wage-earners in 1905 in groups. The last investigation referred to is probably the most reliable *extensive* investigation of wages ever undertaken in the United States. The wages of over three million wage-earners in manufacturing industries are tabulated in a frequency table. Various state bureaus, and the Interstate Commerce Commission as well, give averages and nothing else. The usage in the United States as to the collection of *rates of wages* or *earnings* is not at all uniform.—TRANSLATOR.

[20] This computation is made by dividing the total amount consumed in a country (production plus imports minus exports) by the population.

[21] A start has been made toward statistics of individual consumption by the collection and statistical treatment of family budgets and account books, upon which modern statistics lays great stress.

ments, possible but not actually found, because of obstacles of a statistical character, there are numerous averages for magnitudes which are incapable of individual observation of any sort. Of the latter type of averages are average air space or floor space per inhabitant of a tenement, the average debt of a country per capita, the per capita cost of national and city administration, the per capita expenditures or receipts of the government, the average value or quantity of foreign trade per capita, etc.

In all of these cases individual measurements are inconceivable. It is impossible to measure a definite air space, portion of the government debt, or cost of administration for each individual. In such cases we are not dealing with an "isolated average" for a "potential element of measurement," but with a relative number which originates through interrelating two wholly independent magnitudes. We should not speak of such a ratio as an average, since individual measurements are impossible, but rather as the size-ratio of two magnitudes.

Such size-ratios are very common in all branches of statistics. The most varied masses may be interrelated for special purposes. In each case the number of units in one mass is divided by the numbers of units in the other. Thus the division of the deaths during a year by the mean population gives the death coefficient. If, as is frequently done, the quotient is then multiplied by 100 or 1,000, the relative number indicates how many units of the first mass there are, on an average, to 100 or 1,000 units of the second mass. For example, the death rate (that is, the death coefficient multiplied by 1,000) indicates how many deaths occur on an average to 1,000 living of the mean population.[22] In a similar way, it may be computed

[22] While mortality coefficients as well as death rates originate through interrelating the number of deaths and the mean population, differing from each other only in the position of the decimal point, the probability of death is computed by dividing the number

how many births, marriages, crimes, suicides, etc., occur on an average to every 1,000 or 10,000 living of the mean population. Single values for concrete groups of 1,000 or 10,000 persons each cannot be obtained, since the mean population is an abstraction which cannot be submitted, either wholly or in part, to a constant direct observation.

Likewise the population (or definite groups of the population) and the area are frequently interrelated. The result gives the average population per square mile (density of population). To determine the population of the individual square miles would, of course, be impossible, since these areas are units of computation and not objective values. Similarly the average number of schools, post-offices, etc., the average length of road, railroad, telegraph line, etc., is computed for some round number of square miles. We have also to do with size-ratios when railroad statistics give the freight density (ton-miles per mile of line) or when labor statistics give the average number of applicants for each place offered.

Not only statistical coordinate numbers, but also the equally common subordinate numbers, appear regularly in the form of averages. We say: in 100 infants there are on an average 51 boys and 49 girls; of 1,000 inhabitants, according to the census, so many belong, on an average, to the different nationalities, religions, occupations, etc.; of 1,000 individuals of the same age, so many die, on an average, at specified ages (mortality table). Such subordinate numbers express the composition of larger statistical masses in simplified form by reducing the total to 100 or 1,000. Concrete groups of 100 or 1,000 units each, of course, do not exist.

It is evident, therefore, that those statistical quotients, which refer to virtual elements of measurement are, indeed,

of deaths by the total population among which the deaths occur during the period of years in question and thus the latter differs from the former both in conception and numerical value.

true even though they are "isolated" averages, but that the much more numerous coordinate and subordinate numbers, although they appear as "averages" in statistical language, are not actually averages in the literal sense of the word. Thus the average number of inhabitants per square mile is not an average of the population of *each* square mile of the country, since the determination of such numbers is inconceivable.

But among statistical relative numbers there are some which are not merely nominal averages but are real averages, though, to be sure, in a new sense. We shall now consider these especially interesting relative numbers.

In the preceding section [23] we found that the arithmetic average for certain time, space, qualitative, or quantitative series must be computed as a relative number of a higher order. Such relative numbers are, therefore, true averages with respect to the items referring to parts of the totality. Relative numbers are in their nature averages even though there be no special relative numbers referring to parts of the totality, on the condition, that such special relative numbers are theoretically possible. This condition is frequently fulfilled.

Two examples will make this clear. Statisticians know that the sex-ratio depends upon the density of population of the district (in consequence of selective migration), and upon the age class (in consequence of the varying mortality of the sexes). The general sex-ratio for the total population is, therefore, a mean originating from divergent sex-ratios for different places and age classes, possessing in itself the character of an average, even if the sex-ratios for the various subdivisions are not known. Similarly the annual death rate is an average, since the mortality may fluctuate during the year. This average levels, therefore, the "time frequency" of mortality.[24] It levels, also, the

[23] Cf. p. 18 f.

[24] Compare Mischler's Handbuch der Verwaltungsstatistik, § 30,

mortality differences which usually exist for different geographic divisions, and for different occupations, age classes, and the like.

Suppose we assume that mortality varies with sex, age, and civil condition. A general death rate is the average of " special " death rates for various groups divided according to sex, age, or civil condition. Likewise, the " special" death rates for the groups are themselves averages, as there are numerous other influences affecting mortality. For example, the death rate of individuals of a certain sex, age, and civil condition is an average of the various death rates of the subdivisions of such individuals according to occupation. But there are many other things influencing mortality, such as economic condition, nourishment, housing, the altitude, geological formation, etc. Since a death rate referring to an entirely homogeneous group of the population does not exist, every death rate must always be thought of as an average leveling the divergent rates of component parts of the population in question.

What holds for death rates holds as well for most of the remaining demographic relative numbers, marriage rates, birth rates, and the life. Demographic phenomena show, as a rule, time and space fluctuations which vary in degree among different groups of the population. Age, sex, civil condition, occupation, and economic condition always have their influence. Some phenomena, such as suicide, show the influence of religion; life in cities produces other effects than life in the country. Quetelet sought in his *Social Physics* to ascertain the action of the factors influencing demographic phenomena and numerous other writers of a later date have presented " Schemes of Influences Affecting Mankind," or statistical " Systems of Causes." [25]

" Die Massenerscheingungen als Funktion der Zeit," especially, p. 90 f.

[25] See, for example, in this connection: Ottingen's Moralstatistik,

The fact not always sufficiently regarded, that relative numbers are themselves frequently averages, is of the greatest importance for statistics. Relative numbers should above all make comparisons possible. For this purpose we compare the death rates of different years or classes of the population. But there is, as will be subsequently shown, generally an element of uncertainty in the comparison of averages and hence of relative numbers. Thus the difference in the general death rates of two countries may be caused simply by the different composition of the population of the two countries, while the death rate in the individual population groups or age classes in each country may be the same. A country with a larger number of children will by reason of this very fact show a larger general death rate than a country with a small number of children, even though the death rate in the corresponding age classes of the two countries is the same.

Engel's Bewegung der Bevölkerung im Königreiche Sachsen in den Jahren 1834-1850, Wagner's Gesetzmässigkeit in den scheinbar willkürlichen menschlichen Handlungen, Gabaglio's Storia e Teorica generale della Statistica, and, most recently, Colajanni's Statistica teorica.

Quetelet differentiated between "natural" and "accidental" or "perturbing" influences, the latter originating from man himself. As the "natural" influences affecting mortality he named: the influence of locality, sex, age, the type of year, the seasons, the time of day, and the influence of various diseases; "accidental" or "perturbing" causes of death are: the influence of occupation, economic condition, morality, enlightenment, and the political and religious development. The influences enumerated by Quetelet are more or less generally recognized; their division into the two groups "natural" and "accidental" or "perturbing" has, however, been proven invalid.

Colajanni enumerates, *first*, physical influences, such as climate, soil, moisture, etc.; *second*, anthropological causes, such as sex, age, mental and physical constitution, inherited traits, and race; *third*, social causes, such as density of population, relations brought about by living in groups, divisions of the country, laws, political organization, stage of culture, religion, family position, and occupation.

There are, accordingly, often differences of opinion about the conclusions to be drawn from a comparison of relative numbers. Special methods have been invented for making a comparison of definite relative numbers possible.[26] The most suitable for purposes of comparison are, as will be shown in detail later, those relative numbers which refer to elements containing the least possible differences. These are relative numbers which are computed for most nearly homogeneous masses. Such relative numbers possess, comparatively, the greatest scientific value, and are therefore one of the chief aims of modern statistics. For this reason, those relative numbers which are to be regarded as averages are often resolved into more homogeneous components, which may then be compared, so to speak, as special values with the original relative numbers. Thus special death rates for different age classes, occupations, etc., are being computed and then compared as special values with the general death rate for the whole population. It is then possible in such a case to compare the original relative number with the supplementary special values and in this way to test its scientific value and its applicability for different statistical methods.

B. ESTIMATED ISOLATED AVERAGES

Isolated averages may be estimated as well as computed. In any case the individual items are lacking and the comparison of the average with them is, therefore, impossible. But whereas the computed average is fixed numerically the estimated average is determined between limits of greater or less distance apart. Still, as a matter of fact, the statistician is not infrequently forced to estimate averages because of lack of the necessary data for computation. The

[26] Thus, for example, the method of the standard population serves the purpose of making the death rates of various countries comparable (see p. 159 f.).

estimation of averages for "potential individual observations," and the estimation of relative numbers, which are really averages, are of especial importance.

(a) *Estimation of the Average Size of a Virtual Individual Element of Observation*

In a large number of cases where it is impossible to obtain the individual items in detail and where an isolated arithmetic mean cannot be computed, the problem of estimating that mean, or other mean such as the mode, is intrusted to experts. This procedure is the common one in ascertaining the average wage of agricultural laborers.[27] In Austria and Germany the estimation of the customary wage is necessary for the administration of workingmen's insurance. By "customary wage" is meant the wage received by the greatest number of laborers, that is, the "normal," prevailing, predominant, or modal wage.[28]

[27] See, for example, the collections of agricultural wages in Austria made in 1895 and 1897 by the Landes Kulturräte and Landwirtschaftgesellschaften (Ost. Stat., Vol. XLIV, No. 1, and Stat. Monatsschrift, 1904, p. 466 f.). In many cases the experts have given the maximum and minimum rates as well as the arithmetic mean.

[28] Section 6 of the Austrian laws of March 30, 1888, R. G. Bl. No. 33, concerning sick insurance of workmen, contains the provision that the aid given during sickness should be, at least, "60% of the daily wages customary in the jurisdiction received by laborers entitled to insurance benefits." Section 7 provides: "The amount of the daily wages customary in each jurisdiction and received by laborers entitled to insurance benefits are to be fixed periodically by the political authorities after a hearing of proxies and, in those places where the jurisdictions have representatives, after consultation with committees from the respective jurisdictions. If it is found that wages vary greatly then the customary daily wages of several categories may be established. They may be established for men, women, children and youths, especially, . . ."

In Germany the customary wages received by day laborers are established for the various communities by the higher administrative authorities after giving the local authorities a hearing.

The estimation of averages is also of importance in the railroad business. Since railways transport much freight without weighing each item it is necessary to fix rates upon the basis of estimated average weights. For example, the Austrian railways fix the following average weights: a sucking pig, 20 kg.; a young boar, 30 kg.; a lean hog, 60 kg.; a fat hog, 170 kg., etc.[29]

Many times an estimate of an arithmetic average is made in order that such estimated average may be multiplied by the number of elements to which it refers, thus giving the sum total of the unknown individual elements. Thus we estimate the arithmetic average amount of money taken out of the country by an emigrant in order to ascertain the total amount taken out, which is the product of the average and the number of emigrants.[30] Similarly, various writers have used the following equation in estimating national income:

(Total of incomes reported subject to tax) + (Estimated average of untaxed incomes) (Number) = National Income.

Estimates of the population of the past depend upon preliminary estimates of averages. If we know the number of families, the number of dwellings, or the number of hearths in a city or country and can estimate the average number of persons per family, dwelling, or hearth, the

[29] These arithmetic average weights, which ought to correspond to the actual weights, must be distinguished from the "normal weights" ("Normalgewichte") of railroad tariffs, which differ greatly from the actual weights and are merely assigned to serve in certain railroad tariff computations.

[30] Cf. the Tabellen zur Währungsstatistik issued by the Austrian Minister of Finance, 2nd ed., Pt. II, Vol. III, § 13, "Daten zur Zahlungsbilanz," p. 823, where, for example, it is assumed that the Austrian emigrants to Brazil, Argentine Republic, and Canada take with them, on an average, 100 crowns in ready money. On p. 832 of the same publication it is estimated in a similar way—naturally upon the data from certain stopping places—that the arithmetic average daily expenditures of foreigners stopping in Austria are 15 crowns.

product of corresponding numbers gives the population of the area in question. Methods similar to those just quoted must be used in computing the present population of countries where a census is not taken.[31]

The statistics of the value of imports and exports depend, in most countries, upon estimates of experts. Likewise, estimates of average weights are necessary. In order to express all imported and exported wares as a total, they must be expressed in the same unit of weight. In the case of those articles which are recorded, not by weight, but by the piece or otherwise, it is necessary to estimate the average weight. Thus, in the statistics of the foreign trade of Austria-Hungary, cattle, hats (with certain exceptions), carriages, bicycles, watches, etc., are not reported according to weight but by number. Accordingly, in the year 1904 the average weight of an ox was taken as 450, 650, or 500 kg., depending upon whether it was imported, exported, or shipped in domestic trade.

Many statisticians have also, especially in recent times, tried by means of estimated averages to compensate for lack of statistics of production. Thus Czörnig computed the total production of Austria from the ascertained number of different machines, by ascribing to them a definite average productivity. He says in his preface to the *In-*

[31] See especially the report submitted by Marcus Rubin at the ninth session of the International Statistical Institute (1903) entitled "Sur les exploitations démographiques à exécuter dans les pays, où il n'existe pas encore de recensements." Prof. L. Gumplowicz relates in his Verwaltungslehre that Lord Macartney, a British ambassador, computed the population of a Chinese province from a certain store of salt which he was told was to cover the consumption for one year. Lord Macartney estimated the arithmetic average consumption per capita and then computed the number of persons that could be supplied with the given store of salt at the average consumption estimated. Westergaard (Die Grundzüge der Theorie der Statistik, p. 270) mentions an estimate of population upon the basis of grain consumption made by Crome in 1785 (Grösse und Bevölkerung der europäischen Staaten).

dustriestatistik der österreichischen Monarchie für das Jahr 1856 (p. vi): "In each . . . (of the different branches of industry) . . . there is a technical unit, which serves as the measure of production, as, for instance, the loom in weaving, the machine or the vat in the manufacture of paper, the furnace in the manufacture of steel, the number of workmen in the manufacture of machines, etc., and any expert, knowing this technical unit of an industry, will be able to estimate its production with approximate correctness." This method of computing the production by multiplying the number of machines by their average output, which Block also has recommended in his *Traité de Statistique* (2nd edition, 1886), is, however, for obvious reasons unreliable, and is therefore no longer used.[32]

Frequently, estimates of averages refer to magnitudes which are not only removed from observation in the mass, but which even individually cannot be measured exactly. Thus, various writers have tried to estimate the average capital expended on the education of an adult in different classes of society and hence to express in figures the "value of the individual."[33]

As already stated, various kinds of averages, such as the arithmetic mean, the median, the mode, etc., may be computed from series of individual observations. Hence, not only the estimate of the arithmetic average but also that of the relatively most frequent magnitude may be attempted. But in practice the various kinds of averages are not always sufficiently distinguished; those whose duty it is to make the estimate frequently do not have strict

[32] Lavoisier applied a still more questionable method in 1790 when he utilized the number of plows, found in some way or other, and the estimated average performance of a plow to derive the size of the fields not registered and the amount of the agricultural production of France.

[33] See Dr. E. Engel, Der Wert des Menschen (1883), Pt. I, "Der Kostenwert des Menschen" (with bibliography).

orders and choose the average whose determination is easiest for them. Now it is a special characteristic of the mode that it may be estimated more readily than the arithmetic mean. To determine the latter, all the single cases must be taken into account according to their magnitudes; but if they are known, the average is computed, not estimated. The mode of a set of values, on the other hand, easily impresses the observer, and any expert will be able to give it without computation. It may be surmised, therefore, that estimated averages are oftener modes than arithmetic means.

The arithmetic mean and the mode do not, as a rule, coincide. This fact becomes of great significance as soon as the sum total of all the items for all the single cases is to be computed by means of the estimated average. This sum may be obtained by multiplying the number of cases by the average value of the items, but not by multiplying such number by the mode, unless this latter chances to coincide with the arithmetic mean. For instance, if the average size of a family is multiplied by the number of families the population is obtained, but not if the mode is multiplied by the number of families. Accordingly, estimated averages for a potential element of observation, when used to compute a totality, give a correct result only when they correspond to the arithmetic mean and not to the mode.

Estimated averages, naturally, do not possess the same value as averages computed from detailed data, since those who make the estimate, as a rule, know only a small part of the individual cases and, perhaps, give a judgment from limited observations. Therefore, modern statisticians try, as far as possible, to dispense with estimates of a potential element of observation.

It is to be observed that estimated values occur sometimes without being evident. In such cases there is danger of ascribing to the value in question a greater reliability

than it actually possesses. Thus, for example, the method of direct questioning about individual average wages, often used because of its simplicity, leads generally to mere estimates of averages. If an employer is asked to give the average weekly earnings of his employees, he is generally able to compute them correctly from a series of successive paydays. But it is unlikely that he will take this trouble. He is more apt to attempt a more or less arbitrary estimate.[34]

(b) *Estimation of Relative Numbers which are Themselves Averages*

Relative numbers having the character of averages are also often estimated when there are no sufficient data for their computation. The ratio of two masses is often determined for the purpose of computing the size of an unknown mass from a known one. Statisticians often employ such estimates in order to obtain the population figures of past times. If we know, historically, the number of citizens, or artisans, or slaves of a city or country, we may estimate what percentage of the entire population these various classes probably formed on an average at the time

[34] Individual average wages, in the sense explained above, were asked for at the inquiry of the Reichenberg Handels- und Gewerbekammer in 1888 (see Nordböhmische Arbeiterstatistik, Reichenberg, 1891). In the investigation concerning the conditions of workmen in the Ostrau-Karwin coal district, which was undertaken by the Austrian Bureau of Labor Statistics in 1901 under the direction of Dr. Mataja, the actual earnings of mine workers were determined; arithmetic average wages were collected solely for comparison with the arithmetic average wages collected from workmen of the district employed in other occupations; the question called for the amount each workman received, on an average, per week during the first half of the year 1901. (See Arbeiterverhältnisse im Ostrau-Karwiner Steinkohlenreviere. Published by the k. k. Arbeitsstatistisches Amt im Handelsministerium. Pt. I, "Arbeitszeit, Arbeitsleistungen, Lohn- und Einkommensverhältnisse," Vienna, 1904, p. xlix.)

in question, and hence we may compute the total population.

The ratio between definite statistical masses is often fairly constant and can easily be estimated to be within certain limits. This holds good, for instance, for the relation between population and births or deaths. Since the time of Halley and John Graunt, innumerable attempts have been made to compute the population, lacking a census, by means of estimated death or birth coefficients based on data concerning the movement of population.[35] This method has often been used to reconstruct the population of past times. In Roumania it is still employed.[36] Also in states where questions about religion are not asked in the census, the attempt has often been made to compute the numbers allied with the various denominations from the known number of communicants and the estimated percentage that this latter number bears to the former. In America even the number of seats in the churches is used in making this computation. This estimated ratio of the communicants or seats to the total number in the denomination is just as much an average, since it varies from place to place, as are the general sex-ratio of the population and the general death rate.

Every estimate must be regarded as simply an approximate value. It may be quite accurate in some cases, but there is no certainty. Therefore the number computed by means of an estimated relative number must also be regarded as merely approximate. The fact that the relative numbers in question are averages causes particular difficulties. If, for instance, the population of a country is to be inferred from the number of deaths, the one who makes the

[35] Sonnenfels in his Grundsätze der Politik (Vienna, 1819, Vol. I, p. 32) cites a number of methods of estimating population which are based on the fact that "political computation infers the number of people from relations which are determined by experience."

[36] Compare G. v. Mayr, Bevölkerungsstatistik, p. 15.

estimate usually applies a death rate which has been observed to be true of a certain section of the country or class of the population but which may not hold good for the entire country. Modern statisticians endeavor, therefore, to secure data sufficient to enable them to dispense with estimates, though, to be sure, their attempts have hitherto been only partially successful.

C. SOME INSTANCES OF THE PROGRESS OF STATISTICS AWAY FROM ISOLATED AVERAGES AND TOWARD AVERAGES BASED ON SERIES OF ITEMS (average length of life, length of marriage, length of a generation, number of children per family)

As has been mentioned, the tendency of modern statistics is to obtain not isolated averages but statistical series from detailed observations. From these series averages may be computed whose reliability and cogency may be determined in individual cases by comparing them with the original items of the series. In the following discussion additional remarkable examples will be given of the progress from isolated averages toward averages from statistical series. These examples differ in character from the cases already mentioned and must therefore be treated separately. In the cases which we now take up the chief question is, which of several statistical series furnishes the most valuable average, methodologically speaking. This question may, under certain circumstances, be of great significance, since series of different kinds naturally produce averages of different character; also the value of the average depends, of course, on the individual values contained in the series.

An interesting example of the evolution in question is furnished by the method of computing the *average length of life*. In former days it was generally computed as an isolated average, by dividing the population by the number of births or the number of deaths. But soon doubts began to arise about the correctness of this average. Nowadays

it is a statistical commonplace that the quotient of the population and number of births (or deaths) would give the average length of life correctly only provided that the population is stationary. In such a case one might argue: each birth replaces a living being, in one year the xth part of the population is replaced by new births, in x years, therefore, the total population is so replaced: that is to say, the average length of life is x years. So one might also say of a stationary population: if the yth part of the population is lost by death each year, the whole population must be renewed in y years, and accordingly the average length of life must be y years. In a stationary population, of course, x equals y. But, as a matter of fact, there is no such thing as a stationary population; x and y are always different, and neither of the two methods of computation is theoretically tenable.

Recognizing that population is not stationary, some writers, such as Deparcieux in France and Price in England, have proposed that the total population be divided by the mean of births and deaths. Malthus and Charles Dupin have accepted this method, and Wappäus too has recommended it. Other authors have computed the average length of life as the arithmetic mean of the two ratios: (1) population to births and (2) population to deaths. Both methods are purely empirical but lead to more accurate results than does the mere consideration of *either* the births *or* the deaths.

Statisticians have ceased to compute the average length of life as an isolated average. Instead, they secure data of the length of the lives of the inhabitants of different countries and compute the average length of life as the arithmetic mean of the series thus obtained by observation. Other averages of these series may also be obtained which furnish further standards of vitality. At first, opinions differed as to which series of single observations should be taken to compute the correct average length of life. As

is well known, the arithmetic average age of those living and also the arithmetic average age of those deceased used to be regarded as the average length of life. At present the average length of life or expectation of life at birth is computed as the arithmetic average of all the ages in the mortality table. A mortality table represents the gradual dying at various ages of a number of people born at the same time. It may be based upon a concrete observed totality or, as is the practice, upon an ideal or hypothetical totality, diminished as the ages of the mortality table increase in accordance with the various probabilities of death which are separately determined for the different age classes.[37]

It is manifest that the average age of those living at a definite time (for instance, at a census) does not correspond to the age reached, on the average, by the population. But the expectation of life computed from the mortality table also differs very considerably from the average age of those who have died. A mortality table constructed in the usual way for an ideal totality from recent data pictures present experience. Likewise a mortality table constructed by observation of a concrete group would express the mortality experience of a definite historical period. Such is not the case with the age composition of the deceased and with the average age computed from it. Those who have died during the same period were born in different years. Their age composition is, therefore, caused by the conditions of health which have prevailed at various times, and in computing their average age we combine the effects of independent causes operating at different times. It is now, in

[37] Analogous to these two methods of measuring mortality are the methods of measuring the growth of human beings. Either a certain number of individuals of the same age are observed continuously and measured periodically, or else a number of individuals of different ages are measured at the same time. The latter method is the more practicable, just as it is easier to construct a mortality table for an ideal totality.

fact, generally agreed that only the arithmetic average computed from the mortality table is to be designated as the average length of life, though, to be sure, there are still some open questions as to the way a correct mortality table should be computed. Especially the question of just what groups of living and deceased we should take in determining the correct probabilities of death is one of the most difficult but also one of the most thoroughly discussed chapters of statistics.

The method of computing the *average length of marriage* has undergone a similar evolution. Formerly it was computed by dividing the number of existing married couples by the annual number of marriages contracted or dissolutions of marriage. Given an equal and constant number of annual marriages and dissolutions, the product of this number and of the average length of marriage (expressed in years) would be the number of existing married couples; or, the average length of marriage would be the ratio of the number of existing married couples to the annual number of marriages or dissolutions. But the hypothesis of an equal and constant annual number of marriages and dissolutions does not hold, since the population is not stationary. In an increasing population the number of marriages grows more quickly, that of dissolutions more slowly, than the number of married couples. In such a case, therefore, a division by the number of marriages would make the average length of marriage too short and a division by the number of dissolutions would make it too long. Accordingly, several statisticians have suggested dividing the number of existing married couples by the mean of marriages and dissolutions. Other authors have divided the number of married couples separately by the marriages and dissolutions and taken the mean of the two quotients. Engel, in his work upon *The Movement of Population in Saxony, 1834-1850,* used only the annual number of marriages as a divisor, but corrected it by a coefficient computed for a

long period from the mean annual fluctuation of marriage frequency.

Of course, it was soon realized that these methods were only makeshifts, and that statistics should seek to gain data about the length of all individual marriages and represent them in a series. This has recently been done in several countries. The length of all marriages dissolved by death or divorce has been noted and the average length of marriage has been computed directly from such data just as the average age is computed from the age classification of those who have died during a definite period. The same objections may be made to the former computations as to the latter. In computing the average length of the marriages dissolved during a definite period, we combine marriages dating from different times, which have been exposed to various and independent influences. But it is important to investigate the length of the marriages belonging to the *same* period, which have been affected by more or less similar influences. The most interesting period in this connection is, naturally, the present. The attempt has been made, by comparing the number of couples who have been married a certain length of time with those of the same length which have been dissolved, to compute probabilities of dissolution and thereby to construct a marriage table, analogous to the mortality tables, for a period as close as possible to the present. Such a table represents the progressive diminution from year to year of a hypothetical group of marriages contracted at the same time. From it the average length of marriage may be computed as the arithmetic mean. Other averages may also be determined from the table, as the occasion may require.[38]

[38] Compare R. Boeckh's tables of the duration of marriages in Berlin for the years 1875-6 and 1885-6, in the census report for 1875 (Vol. III, p. 69), or in Bewegung der Bevölkerung der Stadt Berlin, 1869-1878 (Berlin, 1884, p. 78), and in the Statistisches Jahrbuch der Stadt Berlin (Vol. XIV, 1889, p. 30 ff.). From

Wappäus and, subsequently, Haushofer have advocated a different method of computing the average length of marriage. These two writers think it may be obtained by deducting the average at marriage of the two sexes from their average length of life. In this way the average length of marriage is not computed from a series of items; and yet the computation starts from two values which are properly averages of series of single observations. This method is, however, imperfect, because it considers only the marriages dissolved by death, not those dissolved by separation. Moreover, not the general mean length of life but the special mean for those who are married should be used. Even then the result would not be satisfactory, since what we desire is not so much a single general figure for the mean length of marriage as values for different categories of the population, especially age classes.

Besides the length of life and the length of marriage, numerous other phenomena and conditions are measured with regard to their duration. But certain difficulties always attend the determination of the average duration of mass phenomena, primarily because the length of the phenomena cannot be ascertained by a single count or census. Only those cases which exist at the time of the census are considered, and even they are not characterized

these tables the average length of newly contracted marriages and of those of a certain duration could be computed. In order to be able to construct such tables it is of course necessary to find out not only the length of all marriages which have been terminated but also of all which still persist.

Similarly Boeckh has also constructed a "marriage table," based on probabilities of marriage, from which an average age of marriage may be computed, which differs theoretically from the kind usually obtained from the age distribution of those marrying in a certain period of time. (Compare Statistisches Jahrbuch der Stadt Berlin, 1884, p. 14, and Die Bevölkerungs- und Wohnungsaufnahme in der Stadt Berlin for December 1st, 1880, Vol. III, § 2 B, p. 10 ff., Berlin, 1888.)

in the way to be desired. Only the duration of the cases up to the time of the census is given; no conclusion can be made about future duration. Thus a census furnishes merely a somewhat unimportant lower limit. For that reason neither the average length of life nor the average length of marriage can be computed from the data of the census. For the same reason a census of the unemployed, in which the latter are asked about the length of time they have been out of work, does not give a theoretically satisfactory result. The same objection holds when attempts are made, as in Belgium, to get the length of life of industrial enterprises from information concerning the date of their establishment.

In order to be able to determine the duration of mass phenomena, the constant observation of the ending or disappearance of them is necessary. But constant notations are statistically much less practicable than a single census. Hence only a few states or cities note the length of dissolved marriages; data are lacking about the duration of various diseases; general figures about the length of unemployment are wanting; and there is no trustworthy information about the duration of industries, buildings, etc.[39]

Averages may be computed, indeed, directly from the statistical series obtained from constant notations concerning the extinction of the cases in question. But we have just shown in connection with the length of life or of marriage that averages computed from series of immediate observations are, theoretically, not entirely unobjectionable. Such averages are valid neither for the present nor for a definite

[39] Compare Mischler's Handbuch der Verwaltungsstatistik, Vol. I, § 31, "Die Dauer als Eigenschaft der Massenerscheinungen"; and Mischler, "Das Moment der Zeit in der Verwaltungsstatistik," in v. Mayr's Archiv, Vol. I, Pt. I. The statute of the French Office du travail assigns as one of its tasks that of determining the "durée moyenne de l'activité de l'ouvrier dans chaque profession," a task which is hardly soluble, at least in France, with the available statistical data.

past time. Cases which belong to different periods are grouped together simply because they end at the same time. The same objections may also be raised against any series which represents the duration of phenomena ending at the same time, provided those phenomena extend over a considerable period, during which the causes affecting them may have changed.

Theoretically, the best way of dealing with such phenomena is to find the relation between the existing and completed cases of equal duration and thus to obtain probabilities by means of which a table may be constructed. From this table the average duration of the phenomena may be computed and other averages obtained. This method has, as we have indicated, actually been applied in connection with the length of life and the length of marriage. But suggestions only have been made in other directions. Professor Mischler,[40] for instance, speaks of a table of destruction of buildings, whose elements would be formed from the numbers of existing and destroyed buildings and the proofs of their ages. He points out that such a table would form the only correct basis for insurance premiums, rents, the measurement of the periods of exemption from taxation, etc.

Interesting discussions have arisen about the *length of a generation*. G. von Mayr defines it as "the average interval between successive births of the same stock." [41] Here the average duration of definite phenomena or conditions is not involved, but the average period elapsing between certain events. The measurement of this period is of importance in several ways. We may determine by comparing the length of a generation and the length of life what time two or three successive generations live together. "In general, the whole possibility of progressive civilization depends upon the relative time which the different generations

[40] Handbuch der Verwaltungsstatistik, Vol. I, p. 99.

[41] Bevölkerungsstatistik, p. 413.

live together both within and without the lines of descent." [42]

In economics the length of a generation has been used by Foville to compute national wealth on the basis of inheritance taxes.

There are various methods of finding the length of a generation. Rümelin claimed that he obtained it by adding to the average age of men at marriage half the marital period of fertility.[43] Obviously this method leads only to an "isolated" average. Von Inama-Sternegg has suggested computing the length of a generation from mass observations based on genealogies, and he has himself applied this method to a considerable extent.[44] Finally, Vacher and Turquan, taking the data in regard to the age of parents furnished by the birth certificates in France, have computed the length of a generation as the average age of the parents of children born during a certain period.[45-46]

These methods might also be judged from the standpoint assumed previously in the criticism of the methods of measuring the length of life and of marriage. They do not, in fact, give a magnitude which refers to a definite period; on the contrary, they include cases which belong

[42] Georg v. Mayr, loc. cit.

[43] Über den Begriff und die Dauer einer Generation, Reden und Aufsätze, Tübingen, 1875, p. 285 ff. This method, with a slight modification, was also employed by Goehlert; cf. "Die Generationsdauer vom statistischen Standpunkte betrachtet," Stat. Monatsschrift (Vienna), Vol. VII, 1881, p. 49 ff.

[44] "Über Generationsdauer und Generationswechsel" (Comptes-rendus et mémoires, VIII. Congrès international d'Hygiène et de Démographie, 1894, Vol. VII, Budapest).

[45] Cf. § 95 in G. v. Mayr's Bevolkerungsstatistik, "Die Generationen," and also the literature mentioned there.

[46] Quetelet seems to have understood (wrongly) length of generation to mean the average age of men or women at the birth of their first child (cf. Versuch einer Physik der Gesellschaft, German edition, 1838, p. 66 at top).

to different periods. The question might, therefore, well be asked, how the precise length of a generation could be determined for the present or for some definite historical period.

The method of measuring *marital fecundity* offers in its historical development another interesting example of the progress from an isolated mean to one obtained from a series of observations. It is, at the same time, an example of the difficulty of choosing, from several series which apparently represent the phenomenon, that one which will furnish the correct mean.

Marital fecundity may be viewed from two different standpoints. The annual frequency of legitimate births may be investigated. For this purpose the legitimate births of a year are usually placed in relation to the married women living at the same time, either with the total number or only with those of child-bearing age. If greater accuracy is desired, the infants may be divided into groups according to the age of their mothers and may be placed in relation to the numbers of the married women in the age classes in question. Furthermore, the infants may be compared also with the numbers of married men, and finally the various numbers of infants, differentiated according to the ages of their parents, may be compared with the marriages of corresponding age combinations. It was in this latter way that Körösi obtained his "Bigenous Table of Natality," which represents the frequency of births for the various age combinations of parents and which was submitted to the Royal Society of London in 1893, just 200 years after Halley had prepared the first mortality table and presented it to the same society.[47]

[47] See "An Estimate of the Degrees of Legitimate Natality as derived from a Table of Natality compiled by the Author from his Observations made at Budapest" in Vol. CLXXXVI, 2. 1895 B. of the Philosophical Transactions of the Royal Society of London, pp. 781-875. Körösi published a short summary of this article in

It is a different problem, however, to determine how many children on an average result from a marriage during its entire length (or how many are to be expected), in which case, of course, special averages for marriages of different lengths and for different age combinations of the parents are to be sought. Körösi, in the paper cited above, distinguishes therefore between the measurement of fecundity in general and the measurement of "richness of marriages" or "expectation of children." His distinction is of fundamental importance. In the following pages we deal only with the methods by which marital fecundity to death or separation (that is, the average number of children for the entire length of marriage) is determined.

Until fairly recently it was the custom to obtain the life fecundity by dividing the number of births during a year by the number of marriages or by the number of dissolutions for the same year. Körösi quotes this as a method of obtaining marital fertility in the narrower sense. Yet, as a matter of fact, this method, though depending upon a comparison of annual figures, may under definite circumstances give the number of births which occur on an average during the whole length of marriage.

Lexis makes the following statement of the case: if b represents the total number of legitimate births and m the total number of marriages which occur in a generation of females, then evidently the ratio $\frac{b}{m}$ expresses the measure of life fecundity.[48] In a stationary population we could replace b and m (which are not available) by the number of legitimate births and the number of marriages in a definite year and the ratio of the two numbers would express the fecundity. But this is true only if the popu-

1895 in the Revue d'économie politique (Paris) under the title "De la mesure et des lois de la fécondite conjugale." Körösi's Table of Natality was adjusted by Galton and Dr. Blaschke (Vienna).

[48] Cf. Abhandlungen zur Theorie der Bevölkerungs- und Moralstatistik, IV, "Übersicht der demographischen Elemente und ihrer Beziehungen zu Einander," p. 81.

lation is stationary. Even Süssmilch and Malthus pointed out this fact, as Körösi has mentioned in his paper "Zur Erweiterung der Natalitäts- und Fruchtbarkeitsstatistik."[49] Even if the population were stationary, this method of computing marital fecundity would be imperfect, because instead of making a distinction between sterile and fruitful marriages it gives a common average for these two essentially different categories.

But, as is well known, population is not stationary. Therefore, if the measure of fertility is computed according to the method in question, the result will depend chiefly on the number of marriages or dissolutions during the given year. The mere decrease of marriages or dissolutions will result in an increase in the quotient, indicating apparently a greater fertility; and conversely, a smaller measure of fecundity will result from an increase of marriages and dissolutions.

Such a method is, therefore, extremely dubious. Moreover, it is not made correct by dividing the births by the arithmetic mean of the marriages and dissolutions of the same year. Nevertheless, this latter modification of the method has prevailed until very recently. Haushofer as late as 1882 advocated it in his *Lehr- und Handbuch der Statistik* (p. 407).

Two other improvements of the method have been attempted. First, the ratio between births and marriages has been determined for a longer period than a year, so that a larger percentage of the births might come from the marriages used in the comparison than would be the case for a single year. Secondly, not the marriages of the same year (or period) as the births are used, but those

[49] Bulletin de l'Inst. intern. de Stat., Vol. VI, 2nd issue, Supplement 22 f. p. b.; for further discussions about the measurement of fertility and the use of it by various authors, see the same reference, and also R. Boeckh's "Die statistische Messung der ehelichen Fruchtbarkeit" in the Bulletin, Vol. V, 1st issue, p. 159 ff.

of an earlier period, which precedes the period of the births by the average interval between marriage and the birth of a child. Dr. Farr has computed this interval to be six years, and has accordingly divided the number of births by the number of marriages contracted six years before. The divisor is, therefore, somewhat smaller than when the marriages of the same year are taken, where the numer of marriages is increasing, and thus the marital fecundity appears greater. Wappäus has employed the tedious method of comparing the arithmetic mean of the annual legitimate births for three years with the mean of the annually newly contracted marriages (and, if possible, also of the dissolutions) of the seven preceding years.[50]

Modern statisticians try to obtain the average marital fecundity not as an isolated quotient but as an average from series of observations of the number of children in individual marriages. This detailed observation renders possible the distinction between sterile and fruitful marriages. Of fruitful marriages those of different lengths and different age relations of the parents may be differentiated and special averages for the number of children in these categories may be computed.

The first attempts to compute the average number of children per marriage on the basis of detailed observations utilized the results of the census, in so far as questions had been asked about the number of children born in wedlock and about the length of the marriages. But it is evident that such data do not give a measure of the full marital fecundity. The marriages, which the census ascertains, persist and in many of them additional children are born. Such data signify merely a minimum, and give no information about the number of children born during the whole length of marriage.[51] Boeckh has, how-

[50] Allgem. Bevölkerungsstatistik, Pt. II, p. 314.

[51] See the states or cities, for which there are data concerning

ever, attempted to compute marital fecundity on the basis of the census data and of the table of lengths of marriage, presuming that the dissolved marriages of a definite duration might as a rule have the same numbers of children as the persisting marriages of the same duration, which belong to approximately the same period of fertility.[52]

Another method of measuring marital fecundity is to number each birth according to its order in the family in question, and, if possible, also at the same time to ascertain the age of the parents and the length of the marriage. But this method is as unsatisfactory as the one previously described. The marriages investigated are not yet terminated and other children may be born. Furthermore, the percentage of sterile marriages cannot be determined from such lists.[53] In spite of these difficulties, Boeckh in his Berlin statistics undertook, by an ingenious working over of the birth data and with the aid of the mortality table, to deduce the fecundity, just as he had attempted it on the basis of the census data. The results obtained in this way differed considerably from those obtained directly from the data of births.[54]

Only very recently have the marriages dissolved by death or separation been observed with reference to the

existing marriages according to the number of children, in A. N. Kiaer's Statistische Beiträge zur Beleuchtung der ehelichen Fruchtbarkeit, Pt. III, Christiania, 1905 (Tables 1, 3, and 4).

[52] See Die Berliner Volkszählung von 1885, Vol. II, Pt. II, p. 50 ff. March has tried to utilize the results of the French census of 1901 to determine fecundity by considering only marriages of longer duration (more than 15 years, 20 years). Compare his Familles Parisiennes, Composition-Fécondité.

[53] See the states or cities for which we have data of the births according to their ordinal number in Kiaer's Statistische Beiträge zur Beleuchtung der ehelichen Fruchtbarkeit, Christiania, 1905, Pt. III (survey in Table 2).

[54] The results of this computation for the years 1886-1890 and 1891-1895 are given in the Statistisches Jahrbuch der Stadt Berlin für das Jahr 1899, p. 104.

number of children born from them in order to compute the average number of children per marriage from the statistical series thus obtained. This method requires that we should ascertain, in the case of all deaths of married persons and in the case of all separations, the number of children born from the marriages. But since fecundity varies according to length of marriage and according to the age of the parents, it is also desirable that these latter facts should be ascertained.[55]

But even this method of computing marital fecundity is not entirely unobjectionable. The same criticisms may be made against it that were made against the computation of the average length of life from the age classification of the deceased or against the computation of the average length of marriage from the classification of dissolved marriages according to their length. It might be shown that the marriages terminated during a definite period (say, a year) had been contracted during the preceding decades and were therefore subject to varying influences. The mean computed for such marriages is, accordingly, not suitable for characterizing the fecundity of the present time; indeed, it does not apply to any definite period. Professor Zoltán Ráth has formulated this objection in his "*Mémoire sur la méthode la plus simple de mesurer la fécondité des mariages.*"[56] He says: "Although the statistics of marriages dissolved by death include also newly contracted marriages, yet the majority of the marriages whose fecundity is examined date some time back, since in the nature of things the dissolution of marriages often comes after several decades of duration. The method in question, therefore, represents the demographic habits of

[55] Körösi, for example, has obtained such exhaustive data for Budapest. Cf. his "Weitere Beiträge zur Statistik der ehelichen Fruchtbarkeit" in the Bulletin de l'Inst. intern. de Stat., Vol. XIII, Pt. III.

[56] Bulletin de l'Inst. intern. de Stat., Vol. XIII, Pt. II.

past generations rather than of the present." But Professor Ráth is inclined not to lay any great stress on this objection. He says later on: "While admitting that death frequently happens only after a period already sterile in the lives of married women, it may be asserted that in the majority of marriages so dissolved the period of fecundity is not past. The greater the general mortality, the nearer our method will follow the fecundity of the present generation."

At all events, the question arises, in what way the precise average fecundity of the present may be computed. The way seems to be indicated by the somewhat analogous development of the method of computing the average length of life. It was once thought that the average length of life might be found in the average age of the deceased. Now it is computed from a mortality table based on the probabilities of death. Might not the average marital fecundity be similarly computed from probabilities of birth? Ludwig Moser seems to have thought of this. In his work *Die Gesetze der Lebensdauer* (1839), which was epoch-making in its day, he said that in order to determine fecundity we must make our observations in connection with the ages of the parents so as to ascertain what percentage of married women at the age of 30 have a child, what at the age of 31, etc.; the sum of these would give the marital fecundity.[57]

G. von Mayr appears to have a similar method in mind. He asserts in his *Bevölkerungsstatistik* (p. 185) that it would be possible to obtain "a complete picture of marital fecundity in its gradation according to the age conditions of the parents combined with length of marriage (fecundity table)" as follows:[58] "1. Directly and

[57] See Boeckh, "Die statistische Messung der ehelichen Fruchtbarkeit," Bulletin de l'Inst. intern. de Stat., Vol. V, Pt. I, p. 162.

[58] v. Mayr's Table of Fecundity is quite different from Körösi's Natality Table. The latter gives only special birth rates (accord-

strictly historically, by taking a certain group of marriages (say, those of a year) and dividing them according to the ages of the husbands and wives and then ascertaining for each marriage when dissolved the number of children classified according to the duration of the married life. When the last marriage of the group is dissolved, the fecundity of the various groups of marriages, arranged according to age combinations and lengths, is obtained. 2. Indirectly, and ideally, by combining the simultaneous experiences of various groups. A single group of marriages is not observed through the different years of its duration, but fragments of short observations of marriages of different lengths are used to obtain a theoretical fecundity for an ideal group. It is, of course, necessary that we know the age conditions and the length of the existing marriages and that the same facts are known for every birth. If we have all these data the fecundity of all kinds of marriages may be computed, especially such as last until the procreative period ceases.'' The result obtained by the first method of Mayr would refer to a generation already past, in which there would probably be but little interest. His second method, on the other hand, would be analogous to the present prevailing method of computing the average length of life, that is, it would be based on probabilities determined for the present. The latter method, however, has one great fault. It does not express the important difference between fruitful and sterile marriages. Moreover, both methods, so far as they demand a consideration of the length of marriage and of the age conditions of parents, presuppose data which are not obtained at present.[59]

ing to the age classes of the parents), whereas Mayr's Table is intended to represent the development of fecundity in an historical or ideal totality of marriages from contraction to dissolution.

[59] The International Statistical Institute in its 10th and 11th sessions (1905, 1907) took up the question of the formulation of

Körösi is indeed of the opinion that marital fecundity cannot be computed at all on the basis of probabilities of birth. In this connection he says in his paper already referred to, "An Estimate of the Degrees of Legitimate Natality," etc. (p. 867 f.): "One might perhaps think that the addition of the natalities stated for each year of the procreative period would furnish the probability for this whole period. To prove the impracticability of such a proposition it is sufficient to point out the physiological fact that female conception stops not only during childbed, but even during the period of lactation. There exists between two births a natural interval, which, moreover, is further increased by the moral moment. . . . The idea that a wife during five years from the age of 30 to that of 35 could undergo individually the birth probabilities obtained for the total of the wives at the age of 30, 31, 32, 33, and 34 years, is wrong; to observe for one year the natality of five mothers, each of them being one year older than the other, and to observe the natality of one for five subsequent years,—these are two different things." Körösi's arguments, however, do not appear cogent.[60] The single birth probabilities refer to definitely characterized groups of individuals; similarly marital fecundity is to be determined for groups of homogeneous marriages. It may be imagined, therefore, that a definitely characterized group of marriages would be subject successively to the birth probabilities ob-

statistics of fecundity and formed a number of conclusions based on reports of Körösi, March, and Kiaer (cf. Bulletin, Vol. XV, Pt. II, and Vol. XVII).

[60] Körösi's "Natalities" would of course be insufficient, since they are concerned only with the age relationships of the parents but not with the length of marriage. Besides, as v. Bortkiewicz has shown (cf. his discussion of Körösi's Table of Natality in the Jahrbücher für Nationalökonomie und Statistik, 1897, No. 1, p. 123 ff.), these are not really numerical probabilities but "intensity coefficients of marital fertility for individual age classes."

tained for the present time for the age relations and length of marriage in question; and accordingly the total and average number of children might be computed up to the dissolution of marriage on the basis of these probabilities. The birth probabilities also take account of the fact mentioned by Körosi that "female conception stops not only during childbed, but even during the period of lactation," since the women who are temporarily incapable of conception are contained in the denominator of the probability fractions, whence follows a corresponding decrease in the quotient.

CHAPTER III

NECESSITY OF THE LOGICAL AGREEMENT OF MAGNITUDES FROM WHICH AN AVERAGE IS TO BE COMPUTED

Statistical series, from which averages are computed, consist either of quantitative single observations or of values which characterize masses limited in a definite way in regard to their absolute size or in other respects (by relative numbers or averages).

If a series consists of quantitative single observations, these must agree as regards both the observation unit and the observation element in order to produce a clear and precisely definable average. If, for instance, the wages in a definite occupation are to be represented in the form of an average, then only those laborers should be considered who belong to that occupation, and the element of measurement, "wages," must be conceived and obtained in the same manner in connection with all the laborers. The latter would not be the case, if, for instance, both money wages and wages in kind were considered in the case of some laborers and merely money wages in the case of others. The average wage computed from such a series would be neither the average money wage nor the average total wage; indeed, it could not be defined exactly.

If a series does not consist of single observations, but either of values which indicate the size of definitely limited masses (series of the second group) or of values which characterize such masses in other ways by relative numbers or averages, then the individual values do not agree completely, but differ from each other according to a criterion

of time, place, quantity, or quality. But this is the only difference which may exist between the individual values, so that these values must be like several species of the same genus. When the average is computed, the criterion used to differentiate the items is disregarded, so that these items agree completely at that moment. If, for instance, we have the numbers of births for a series of years, these values are, to be sure, differentiated in regard to time, but must agree in all other respects, which would not be the case if, for example, in some years all the births, in other years only the living births, were taken. In computing the average number of births per year the time-difference of the items is ignored and the mean size of masses, each corresponding to the same generic conception, is computed. Relative numbers, especially, which refer to definitely limited masses, must agree exactly in the way in which they characterize the masses to which they refer. It would accordingly be incorrect to obtain an average from birth rates which included the stillborn for some years but not for others.

With certain reservations, therefore, the necessity of a logical agreement of the magnitudes from which an average is to be computed may be postulated. In the case of single observations this necessity is absolute. In series of the second and third groups the items must correspond to the same generic conception, and therefore must logically agree, at least at the moment of computing the mean, at which time the criterion used to differentiate the parts is ignored.

This logical agreement of magnitudes from which an average is to be computed is by no means always satisfied. The agreement of the items of a series is not always easy to establish, especially in time and geographical series where the members originate from distinct investigations. In different returns the same object may easily be differently defined and limited. If in two successive censuses

the idea of a "household" were differently conceived, it would hardly be permissible to compute from these two censuses an average of "households"; such an average could not be clearly defined. Similarly the annual sums of exports and imports would not be logically corresponding magnitudes, if at different times varying quantities of wares were excepted, or if the methods of determining the values had changed considerably. In series obtained from a single investigation it is of course assumed that the conception of the object does not vary. Mistaken interpretations may be regarded as accidental errors, which do not invalidate the logical unity of the series.

The rule that averages must be obtained from logically agreeing magnitudes was formerly frequently violated. We shall not speak of errors due to extreme carelessness. Those cases deserve mention, however, in which magnitudes of different kinds were consciously but mistakenly employed to compute averages. Thus, as has already been mentioned, the attempt used to be made to obtain the average length of life either by dividing the population on the one hand by the number of births and on the other hand by the number of deaths and then computing an average from the two manifestly unlike quotients, or by dividing the population directly by an average of births and deaths —an average of two entirely heterogeneous magnitudes. In the same way it used to be thought that the average length of marriage was obtained by dividing the number of existing marriages on the one hand by the dissolved, and on the other hand by the newly contracted marriages, and determining the average of the two quotients, or else by dividing the number of existing marriages by the average of the dissolved and the newly contracted marriages. Such computations had no scientific basis; they were mere makeshifts. Two possibilities were seen of computing the average length of life; in both methods the dividend was the same (namely, the population), while there were

two divisors, the number of births and the number of deaths. It was found that by using one divisor too large a figure was obtained, by using the other, too small a figure. Accordingly, the simple device was adopted of taking the average of the quotients or of the divisors.

Formerly, on account of the undeveloped condition of statistics, it happened frequently that several methods would be employed, not one of which was felt to be perfectly reliable, and then an average would be taken of the results of the different methods. The political arithmeticians often did this; Petty, for example, computed the period of doubling of the population by different methods and then took an average of his results.[61] Laymen often have recourse to these makeshifts. Leibnitz relates in his *Nouveaux Essais sur l'entendement humain* (1700) that in Lower Saxony when a piece of property was to be sold the peasants used to form three groups, each of which made an estimate, and then the average of the three estimates was fixed as the price.[62] Often, when it was discovered that mortality tables computed in different ways gave different results, averages were taken from several tables. Westergaard remarks in this connection:[63] "Such an employment of several tables, as if they were all equally good observations, is now rightly regarded as irrational."

This process of taking an average of results obtained in different ways reminds one in certain respects of the repeated observations of the same object, so common in astronomy and geodesy, and of the "objective" means computed from those observations. In neither case is it a question of obtaining an average for a series of con-

[61] Cf. Westergaard, Die Grundzüge der Theorie der Statistik, p. 255.

[62] Mentioned in the Journal of the Royal Statistical Society, Vol. LIV, 1891, p. 451.

[63] Die Lehre von der Mortalität und Morbilität, p. 106.

crete phenomena of different sizes, but of the correct ascertainment of a single fact, of the most probable value of a magnitude, for which by reason of insufficient methods of observation several values are suggested. However, there is an important difference. In astronomy and geodesy we have to deal with manifold measurements, undertaken according to the same method and with the same instruments, subject only to accidental errors which are to be eliminated as far as possible by the computation of an average. On the other hand, when, for example, an average is computed from several mortality tables, we are treating as equivalent results obtained by means of different methods and therefore showing definite variations. But it is the duty of science in such a case to ascertain the most correct method and to accept the result of it; it is an abdication of the scientific method to regard results of different methods as of equal value and to blend them in an average.

CHAPTER IV

POSTULATE OF THE GREATEST POSSIBLE HOMOGENEITY OF SERIES FROM WHICH AVERAGES ARE COMPUTED, AND OF MASSES WHICH ARE CHARACTERIZED BY RELATIVE NUMBERS

The significance of averages consists in the fact that they express the result of the activity of definite complexes of causes in one characteristic figure. Thus, the average wage of a definite group of laborers gives a measure of the factors determining the level of wages in that group. Now it is of great importance that the average shall refer to a complex of causes as nearly unified as possible, since only in this way will it possess a definitely intelligible content, and only in this way may reliable inferences be drawn from a change in the average.[64] If masses of items, which have evidently been variously influenced by quite independent causes, are taken together in a series the average so computed has little scientific value, since it does not express the activity of a unified complex of natural or social causes and is, as a rule, poorly adapted to purposes of comparison. If the wages in two different branches of industry are determined by quite different causes, then the average wages for all the laborers in both industries cannot be regarded as a measure of the factors operative in either of them, and hence no trustworthy inference may be drawn from a change in the average. For these reasons modern statisticians try to form series of individual values as nearly homogeneous as possible and to compute averages only from such series. This is true both of series of single

[64] Cf. below, p. 101 f.

observations and also of series whose members indicate the magnitude of definitely limited masses (parts of a larger totality). For similar reasons, in computing relative numbers, the general object is to distinguish masses as nearly homogeneous as possible, and to characterize them by special relative numbers.

A. POSTULATE OF THE GREATEST POSSIBLE HOMOGENEITY OF SINGLE OBSERVATIONS FROM WHICH AN AVERAGE IS COMPUTED

There are two problems to be distinguished in following the tendency to compute averages from series of single observations as nearly homogeneous as possible. The first problem is to eliminate such individual cases as do not show the observation element in question; the second problem is to divide into masses as nearly homogeneous as possible the individual cases which actually do show the observation element.

First problem. In the statistical observation of a number of items it is often evident that a part of them do not show the measurement or other observation element at all. The question now arises, whether in such circumstances the whole series is to be considered in computing the average or only those cases where observation has yielded a positive result. This question is, indeed, unimportant for the determination of the mode, but may seriously affect the arithmetic mean or the median.

The question cannot be answered in general. In those items which do not show the observation element, the causal complex which produces it in other items may not have been operative at all. In such a case the inclusion of the items not showing the observation element would influence the result wrongly. The average size of the observation element would then appear smaller than would be the case if only positive results were taken; or, in other words, it

would depend essentially on the proportion of observations with and without a positive value. On the other hand, if there is no essential difference in the causation of the two classes of observations there is no reason to eliminate the latter.

The elimination of the items which do not show the observation element is often a matter of course, but in other cases the decision of the question is very difficult. In computing average wages, for example, cases where laborers for personal reasons receive no wages at all will naturally be disregarded. More difficult, on the other hand, is the question of dealing with sterile marriages when the average number of children per marriage is to be computed. Generally, a common average for fruitful and unfruitful marriages is computed. But more recently the demand has arisen to disregard sterile marriages completely and to consider only those marriages in which there is at least one child. Of course, in that case the average will be considerably higher. The reason alleged is that absolute unfruitfulness of a marriage is normally a pathological phenomenon and is to be attributed to disability or sickness in one or both of the parents. The causes influencing the degree of fecundity, it is said, are not operative in such unfruitful marriages, and a correct expression of these causes can only be obtained by eliminating wholly unfruitful marriages in the computation of an average. But these arguments are not convincing. Not all unfruitful marriages should be attributed to pathological causes. It is probable that "neo-Malthusianism," which generally affects the degree of fecundity by limiting the number of children, sometimes leads to complete abandonment of reproduction. In that case, neo-Malthusianism would represent a cause affecting both fertile and sterile marriages. The same is also true of diseases which affect reproduction. If such a disease is present at the outset, the marriage will remain unfruitful; but if it appears

during the marriage after children have already been born, it will affect merely the number of children or the degree of fecundity of the marriage. It cannot be asserted, therefore, that the causes of unfruitful marriages are quite independent of the causes which determine the degree of fecundity of fruitful marriages. However, a separate treatment of unfruitful marriages is no doubt necessary. The percentage of unfruitful marriages is of the greatest interest. Average marital fecundity is to be given, if possible, both inclusive and exclusive of them.[65-66]

Of course, to disregard unfruitful marriages must be considered theoretically as a makeshift. The aim is to keep apart items influenced by different causes. As a matter of fact, the items in question are differentiated according to whether a definite effect has been produced or not. The more correct proceeding (although of course not feasible in the present instance) would be to differentiate the items according to the criterion to which the non-appearance of the observation element is attributable. If, then, we assume that unfruitful marriages are, as a rule, to be attributed to pathological causes, we should not eliminate the unfruitful marriages but those marriages in which

[65] Dr. Friedrich Prinzing distinguishes in his Handbuch der medizinischen Statistik (p. 31) between childless marriages in which no child capable of living is born and sterile marriage in which not even a miscarriage has taken place. Practical statistics cannot make this fine distinction but regards all marriages as childless or sterile in which there is no offspring either living or stillborn. Besides Prinzing, A. N. Kiaer (in Pts. I and II of his Statistische Beiträge zur Beleuchtung der ehelichen Fruchtbarkeit, Christiania, 1903) has offered an exhaustive comparative presentation of childless marriages.

[66] For similar reasons it is also desirable that the average length of life should be computed both for all births (including stillborn) and separately for those born alive. According to the German mortality table the average length of life for the latter is 35.58, for the former 34.04 years.

pathological causes occur, and thus the homogeneity of the series would be established.

The elimination of those items which do not show the observation element is, of course, only possible on condition that individual observations have been made. But since such individual observations are not made, for instance, in regard to the consumption of alcohol or meat for the whole population, manifestly those persons or classes who do not consume alcohol or meat cannot be eliminated. The average consumption of alcohol or meat per head for the entire population may indeed be computed as an "isolated average," but no average which applies only to the consumers can be obtained, and it remains unknown whether the whole population or only a fraction of it takes part in the consumption in question.

General averages of the kind just indicated do, however, possess a certain scientific value, even when it is possible to compute special averages for the classes of individuals actually concerned. Such averages give a measure of the significance which the phenomena possess for a wider though indirectly interested circle of people. In fact it is customary in different branches of statistics, even where individual data are at hand, to compute not only special averages for those personally concerned, but also general averages for a wider circle, including also those indirectly interested. Thus, in connection with sick funds, not only the average number of days per case of sickness is computed, but also the average number of days for each member of the organization, in which case account is taken of members who have not been sick at all, since they too are indirectly affected by the duration of the illnesses, the amount of their contributions depending upon it. Similarly, both the average taxes per taxpayer and per head of the population are computed, the latter, so to speak, as a measure of the burden on the whole population. In these and similar cases the absence of the element of

measurement (length of sickness, amount of taxes) must not be attributed to radical differences in causation. The same causes which diminish the size of the element of measurement may by operating more intensively cause it to disappear entirely. Hence it is permissible and, from a certain point of view, instructive to take together all the items, including those which do not show the element of observation.

Second problem. The endeavor to obtain homogeneous series of single observations is only partly satisfied by eliminating cases which do not show the element of observation. Among those cases which have yielded positive data, special parts may often be distinguished which are influenced by different and independent causes. The statistician must try in such cases to divide the whole series into more homogeneous parts. If this is not done, the average computed from the whole series is not the expression of a unified complex of causes, and so does not allow any reliable inferences to be based upon it. Its magnitude will depend chiefly on the proportion in which the various more homogeneous parts forming the series stand to one another. For instance, let us suppose that a series of wage data includes the wages of all laborers in a district. Now it is well known that sex has a determining influence on wages. Women generally receive lower wages. We have, therefore, first of all to separate men and women. If this is not done, the value of the average wages of men and women together will depend on the proportion in which the two sexes are represented, without furnishing any information about the wage conditions of either sex by itself.

But sex is by no means the only factor influencing wages which statistics can determine. Laborers must also be distinguished in regard to occupation, category of work, age, etc. In this way the statistician will form more homogeneous groups and compute special averages for

them. It should also be mentioned here, that series of heterogeneous parts generally show an irregular formation —often with several points of concentration—but that by disintegrating such series regular constituent series are often obtained which furnish a " typical " mean.

In fact, in all branches of statistics there is a noteworthy tendency to form more detailed homogeneous masses. Not only wages but also various other observations are differentiated according to sex, conjugal condition and age, wherever feasible. Distinctions of occupation, economic condition, etc., are sought. Still further divisions may be suggested by the nature of the investigation; for example, in connection with marital fertility, marriages may be divided according to their length and to the age of the contracting parties.[67]

There are, however, many difficulties to be overcome in forming homogeneous masses. In particular, the differentiation of the mean length of life for various classes of the population (according to occupation, wealth, dwelling conditions, etc.) is still very incomplete; and the same is true of averages representing age at marriage, marital fecundity, etc. Many series cannot be divided into

[67] The principle of forming the most homogeneous groups possible is also to be applied to estimates. In Austria the political authorities ascertain the daily wages in the various judicial districts of all workmen who come under the sickness-insurance law. "If considerable divergencies are shown, the ordinary wages may be expressed in several categories. Separate statements are made for male, female, juvenile, and adult laborers. Apprentices, assistants, unsalaried clerks, and others who draw small wages or none at all are classified among the juvenile laborers" (§ 7, Krankenversicherungsgesetz). The decree of the Ministry of the Interior of January 20, 1894, says, moreover, that in case these distinctions are insufficient, other categories may be formed, especially of male laborers drawing full wages—for instance, into foremen, artisans, factory employees, and ordinary day laborers. Furthermore it may be often necessary to make distinctions among the various groups of industries.

homogeneous groups at all, although their structure shows that they are made up of heterogeneous components.

The postulate of homogeneity is not limited to criteria of quantity and quality. Homogeneous groups of observations are also to be sought in connection with criteria of time and place. It is a manifest disadvantage that an average should be computed from items belonging to widely different periods. Of particular significance is the differentiation according to abstract time criteria; the most important of these are the seasons, which, as is well known, exert a considerable influence on demographic and economic phenomena. It is also undesirable to treat very large geographical districts as units. The average in this way loses reality, and becomes a mere abstraction. Here too a differentiation according to abstract geographic criteria is possible—altitude, soil, climate, etc. Especially a differentiation of the length of life according to such criteria is much sought after.

B. POSTULATE OF THE GREATEST POSSIBLE HOMOGENEITY IN THE COMPUTATION OF AN AVERAGE FROM VALUES WHICH EXPRESS THE SIZES OF MASSES LIMITED IN A DEFINITE WAY (CONSTITUENTS OF A GREATER TOTALITY)

Series of the second group which do not consist of single observations but which express the size of definitely limited masses, may also be tested in regard to the homogeneity of the items. Time series of absolute figures are especially to be considered here. Just as there are single observations with the numerical value of zero, so the item zero may also occur in a time series. Such a member may be disregarded in the computation of the average if it was subject to abnormal time influences. But such cases occur only rarely. It will frequently be possible, on the other hand, to keep apart periods of time in which essentially different causes were operative, and to characterize them

by special averages. Accordingly, separate averages ought, if possible, to be computed for the years preceding a new cause and for those which follow.

When an average is computed from time series, the maximum and minimum of the series are often disregarded. This is done from an intuitive effort after homogeneity. It is assumed that the extreme cases have been influenced by special, transitory causes, and that accordingly such abnormal cases should not be considered if an average is to be obtained representing probable future development. But the maximum and minimum need not always be abnormal cases arising from exceptional conditions. And, on the other hand, there may be other abnormal cases besides the maximum and minimum. Therefore, when an average is computed from a time series, it is only permissible to disregard certain years (no matter how or to what extent they differ from the average) when they are demonstrably subject to exceptional causes which supplant the normal causes.[68]

C. POSTULATE OF THE GREATEST POSSIBLE HOMOGENEITY OF MASSES WHICH ARE CHARACTERIZED BY RELATIVE NUMBERS

Relative numbers, which are to be regarded as averages, do not arise by computation from series of individual values, but are obtained by independent subdivision or coordination of statistical masses. Therefore, the postulate that only homogeneous individual values should be considered must be somewhat modified in connection with relative numbers. The demand should be made here that masses as homo-

[68] Single years may also be disregarded for definite non-statistical reasons. Thus various Austrian railroad concessions state that in order to determine the purchase price the seven years of operation preceding the purchase are taken and of these the two most unfavorable years are eliminated and then the average net profits for the remaining five years are computed.

geneous as possible be distinguished and characterized by special relative numbers (subordinate or coordinate numbers). Accordingly, masses not participating in the phenomenon in question are often eliminated, and even masses with positive measurements are generally divided into more homogeneous parts.

Elimination of non-participating masses in the computation of relative numbers. In computing subordinate numbers it may be necessary to eliminate masses which by their very nature belong to one definite subdivision of the whole and cannot belong to any other. Thus, for instance, children are only unmarried, never married or separated. Hence the division of the whole population according to family condition is of doubtful value; the ratio of the married to the unmarried depends essentially on the age division of the population; a large number of children naturally increases the percentage of the unmarried. It is, therefore, more to the purpose to eliminate that portion of the population not yet capable of marriage.

Of greater importance is the elimination of non-participating masses in computing various demographic coordinate numbers. These are often computed by relating definite events and the entire population. But in the whole population constituent masses may often be distinguished in which such events cannot occur. Thus it is impossible for children to contribute to the number of births or marriages. If a measure of the causes determining the frequency of marriage or birth is to be obtained, it is necessary to eliminate from the divisor those classes in which these causes are not operative. This is done by relating the births only to those age classes of the population capable of reproduction, the marriages only to people of marriageable age. In this way not general but specific figures are secured.

Yet some significance must be conceded to the general figures, since they represent the phenomenon in question

from the standpoint of the total population who are at least indirectly interested. G. von Mayr has defended general frequency figures from this point of view.[69] He notes that they are often looked upon with scorn, and goes on to say: "This scorn is justified, so far as we have in mind the question of the subjective participation of the population in the events. But it is not justified if we observe further that not only the factor of subjective participation or responsibility is important, but also the objective burden (in the good and bad senses) of the whole population. Crimes have a statistical interest not only from the standpoint of the subjective participation of individual classes of the population, but also as an objective disturbance of the whole community. Births and deaths may also be so regarded. They are not only interesting as indicating the contribution of the classes capable of reproduction and as indicating the dangers to life of the different classes but they are also significant socially and economically in their relation to the total population."

Division of masses into more homogeneous parts. The elimination of non-participating masses is frequently insufficient. Within the participating masses, constituents may often be distinguished in which a very definite character prevails, or in which the phenomenon appears with different intensity. Where this is to be attributed to different causes influencing the whole part in question, it is obviously desirable to keep such part separate and compute for it special relative numbers.

As an example of the disintegration of a mass into more homogeneous parts, let us take again the division of the population according to conjugal condition. If the composition of the whole marriageable population is represented according to conjugal condition (single, married,

[69] "Die statistischen Gesetze." Public lecture of August 27, 1895 (Bulletin de l'Inst. intern. de Stat., Vol. IX, Pt. II, p. 304 f.).

separated, etc.), no homogeneous mass lies at the basis of the subordinate numbers. Such a division affects men and women differently; it is well known that there are more widows than widowers because of the different ages of the sexes at marriage. Such a division also varies for different ages. As recent investigations have shown, it likewise varies in different social classes; the possibilities of marriage of industrial laborers differ from those of the agricultural population. Possibilities of marriage also depend on economic conditions, etc. It is, therefore, of great interest to determine the division according to conjugal condition not only for the whole marriageable population but also for the more homogeneous parts of it.

More important are the cases in which, in the computation of *coordinate* numbers, the masses to be related may be divided into more homogeneous parts. The differentiation of the demographical frequency numbers, such as birth rate, death rate, marriage rate, etc., is to be considered in this connection. The criteria to be applied are the same as are used to obtain more homogeneous subordinate numbers or to divide series of single observations into more homogeneous groups. Sex, age, and conjugal condition are most commonly used. But occupation, dwelling, sanatory conditions, etc., should also be employed.

The tendency of modern statistics is everywhere to obtain relative numbers for masses as nearly homogeneous as possible. In this process, relative numbers which are by nature averages are often resolved into more special values, which must indeed still be designated as averages, but which are related to the original relative number as individual values for smaller parts. This process evidently increases in importance with the increase of grades of intensity included in the original relative number.

Not only quantitative and qualitative homogeneity but also unity in regard to matters of place and time are to be sought in relative numbers. Homogeneity in place leads

to the formation of "natural districts," which can be marked off according to topographic, hydrographic, and other points of view ("geographic method" in von Mayr's terminology in contradistinction to the "statistical geographic method," by which the geographic distribution of the various grades of a phenomenon is represented). In order to obtain homogeneity in matters of time, Professor Mischler has suggested the formation of "natural time-periods,"[70] differing from the usual calendar divisions and combining periods which exhibit the same frequency of events.

D. SIMPLE INDICATION OF THE RANGE OF A SERIES

In accordance with the principle that averages should be computed as far as possible from series whose items may be regarded as homogeneous, the average is often not found for heterogeneous series. So, too, the computation of a relative number as an average of certain special values is often abandoned, when this relative number would have to be computed on the basis of very heterogeneous masses. When the computation of the average is found inadvisable, there remains nothing to do but enumerate all the individual values or form magnitude classes from them. This, however, does not result in such simplification and brevity as the computation of an average would give. Therefore, the minimum and maximum of the series, the so-called "range" of the series, are often substituted.[71] For in-

[70] Handbuch der Verwaltungsstatistik, Vol. I, p. 89; cf. also by the same author "Das Moment der Zeit in der Verwaltungsstatistik" in v. Mayr's Allg. Stat. Archiv, Vol. I, Pt. I.

[71] This procedure is often chosen without reference to the homogeneity of the series or of the masses in question because it requires no work and yet is sufficient for certain purposes.

This method of characterizing a series by simply mentioning its

stance, frequently only the highest and lowest prices of a commodity are given, when the computation of an average price would be valueless because of considerable qualitative differences. Stock quotations also are generally given only in highest and lowest prices.

This last process is often employed in order to express briefly geographical series of relative numbers which refer to various countries. Thus, the general birth rate for different countries for the years 1887-1891 may be said to have varied from 22.8 (Ireland) to 42.8 (Hungary).[72] Apart from technical difficulties, it would hardly be permissible to compute averages from such geographical series, since a wholly untypical value would result. Even within the same country the conditions may vary to such an extent that a general average would appear useless. In connection with the agricultural wages in Austria in 1893, von Inama-Sternegg refused to compute averages for larger districts from the average wages for the various court-districts given by agricultural experts. "The value of these court-district data consists in their reality; averages for larger districts or for whole countries would obliterate all characteristic differences; the larger the territory, the farther removed the average would be from reality, without other compensating advantages aside from securing a formal, unified expression. On the other hand, it seems useful

highest and lowest values is to be distinguished from the case where in securing data only the maximum and minimum of a phenomenon (for example, the highest and lowest prices or wages) are obtained. Thus the Austrian Labor Statistical office, in securing data concerning the miners in a certain district, also obtained data in regard to the agricultural industry of the same district. The wages of the agricultural laborers were in part ascertained only in maximum and minimum figures. (Cf. Arbeiterverhältnisse im Ostrau-Karwiner Steinkohlenreviere, Dargestellt vom k. k. Arbeitsstatistischen Amt im Handelsministerium, Pt. I, p. xxv and p. 577 ff.)

[72] Cf. the figures for the individual countries in v. Mayr's Bevölkerungsstatistik, p. 177.

to record the maximum and minimum values found for the larger districts.''[73]

[73] The agricultural wages in the kingdoms and provinces represented in the Reichsrat according to special data collected by the Ministry of Agriculture for the year 1893. Elaborated by the bureau of the Statistical Central Commission in Vienna. Österreichische Statistik, Vol. XLIV, Pt. I, Vienna, 1896, p. ix.

CHAPTER V

FORMATION OF MAGNITUDE CLASSES

Statistics can only rarely measure all the degrees of magnitude and intensity which natural and social phenomena display. Frequently, the exact measurement of the items is not attempted; the statistician merely determines to what magnitude classes they belong. Even when the items are measured exactly, the individual observations are, as a rule, finally united in magnitude classes, and even those statistical series, which do not consist of single observations but of values of other kinds, are often treated similarly. This method shows certain analogies with the method of computing averages. The same tendency toward simplification and abbreviation by the omission of unimportant details leads first to the formation of magnitude classes and finally also to the computation of averages. It will, therefore, be advantageous to discuss briefly the formation of magnitude classes before discussing the purpose of averages. This order is also necessary because the statistical series, from which an average is to be computed, frequently consist not of single observations but of magnitude classes, and this fact must be especially considered in the computation of the average.[74]

Frequently, as we have stated, items are not accurately

[74] Cf. for the questions connected with magnitude classes:—G. v. Mayr, Theoretische Statistik, § 43, "Die Zusammenzüge," and § 45, "Verhältnisberechnungen" (p. 94); A. Bertillon, "La théorie des moyennes en statistique" (Journal de la Société de Statistique de Paris, 1876, p. 302); G. Th. Fechner, Kollektivmasslehre, VII, "Primäre Verteilungstafeln" (§§ 47-52), and VIII, "Reduzierte Verteilungstafeln" (§§ 53-67).

measured, but merely assigned to a magnitude class. If a tabular form is used for recording the observations, the number of magnitude classes is *a priori* limited and nothing is noted except the number of observations falling into each class. But even individual records do not always imply exact measurements. Thus, years are often asked for where months and days would be possible; in anthropometric measurements one is frequently satisfied to determine the size to the nearest centimeter. In such cases individual differences which do not reach a certain amount are not expressed, and items which apparently coincide might differ if greater accuracy were observed. The units of measurement that are distinguished (in the above examples, the years or centimeters) are really magnitude classes, and the observations are made with sole regard to these classes.

Certain " continuous " phenomena can never be measured with complete accuracy, so that the resulting data always possess the character of magnitude classes. For instance, space, time, and weight measurements usually concern continuous phenomena. If the phenomenon is discontinuous, consisting in the individual case of indivisible, concrete units, then all the items can be ascertained with complete accuracy. But such accuracy is not, as a rule, necessary.[75]

The preparation and publication of material usually leads

[75] In forming magnitude classes a distinction must be made between continuous and discontinuous data. The latter, consisting of units not further divisible, occur, for example, where human beings of a definite category are concerned. Links between such units are inconceivable. For instance, if industries are classified according to the number of their laborers, one class will end with 49 laborers and the next begin with 50. On the other hand, where measurements are concerned such a method of delimiting the classes is impossible. There may be innumerable links between 49 and 50 cm. It is therefore more accurate to quote a single figure as a limit, so that one class may reach to 50 cm. inclusive, and the next begin with over 50.

to still further leveling of details. For instance, even where accurate age data are obtained, in publication only years are given, perhaps only age classes of several years each. The number of laborers in industrial enterprises, or the wages of laborers, or incomes, or the like, are generally published only according to magnitude class. Individual differences, which are smaller than the range of the magnitude classes concerned, cannot then be determined. Series of quantitative single observations consist, accordingly, either of the original individual magnitudes expressed with more or less accuracy, or of magnitude classes. The items belonging to the different magnitude classes may be given in absolute numbers or as percentages of the total number of items.[76]

Similarly, series of the second group, namely, those whose members indicate the magnitude of definitely limited masses, may be grouped in magnitude classes. Thus, the districts of a country may be united in several groups according to their population. The same is true of series of the third group, those which consist of relative numbers or averages, by which definitely limited masses are characterized otherwise than as regards their magnitude. Instead of giving the density of population or the average wages for each single district of a country, magnitude classes may be formed and the number of the districts in each class given. Time, place, and qualitative series are changed of course,

[76] Fechner (Kollektivmasslehre, §§ 47-67) divides measurement data into the "first lists," on the one hand, and "primary" and "reduced tables," on the other. In the "first list" the measurements are given in the accidental order in which they were obtained. When these measurements are arranged according to size, a "primary table" is effected. By grouping the measurements a "reduced table" is obtained. In practical statistics, however, there is no necessity of constructing a "primary table" when magnitude classes are formed whose limits are predetermined. In such cases the individual measurements are simply assigned to the classes, and do not need to be arranged in them according to size.

when magnitude classes are formed, into quantitative series.[77]

Magnitude classes are scientifically justifiable. By the formation of groups the original data, often very extensive, are compressed into a series which can more easily be surveyed and judged. In well chosen magnitude classes the structure of the mass finds clear expression; regularities are revealed which might not otherwise be discovered. It is also, naturally, much easier to compare two or more simplified series of magnitude classes than series which give the full details of the original material. Of course, in the individual case everything depends on the kind of magnitude classes formed. The object should not be to form as many classes as possible but to form such as express what is characteristic in the structure of the mass. Both the extent and the position of the magnitude classes are of consequence; for magnitude classes of equal extent may, according to their positions, reveal different characteristics.

It is also worth noting that in correctly formed classes a large part of the errors which affect individual observations disappear. A good example of this is the age classification of the inhabitants of those countries in which the age, not the date of birth, is asked in the census. If each age year is given separately, it will be found as a rule that those ending in 5 or 0 are overstated, while the adjacent ones are understated. But if the age years are united in groups of five, for instance, so that the years with round numbers come in the middle of the group, the errors counterbalance each other.[77a]

[77] Just as quantitative observation data are arranged in magnitude classes, so qualitative data are classified into groups with peculiar characteristics. Thus, in statistics of occupation certain vocations, etc., constitute the classes; in trade statistics the various wares are so arranged. But since qualitative data cannot furnish an average, such groups need not be considered further.

[77a] Cf. "A Discussion of Age Statistics," by A. A. Young in Census Bulletin No. 13, and also "The Age Returns of the Twelfth Census,"

To be sure, it is possible to go too far in the formation of groups. The characteristic features of the material may be lost in the leveling of details. Unfortunately, the formation of groups does not depend solely on methodological considerations. The more the statistics are simplified, the less the cost of preparation and publication. Thus, financial considerations are often of influence, which is especially the case with official statistics.

For all these reasons, it will be readily understood that statistical masses, especially such as deal with single observations, are normally expressed in magnitude classes and only exceptionally in full detail. The limited amount of space available also makes it impossible to publish all the individual data about such matters as income, wages, etc. But if a mass is represented in magnitude classes, only as many frequencies, or numbers of items, need be given as there are classes, with of course a statement of the limits of each class.

The formation of magnitude classes may proceed according to various principles. First, classes may be formed whose limits are determined previously (without reference to the number of items falling in each). In such cases, the extent or width of the classes is also previously determined. Generally the same extent is chosen for each class; but classes may also be formed of unequal extent, following some natural formation.[78] Secondly, classes may be formed to contain the same number of items; in these

by W. B. Bailey and J. H. Parmelee, "The Census Age Question," by A. A. Young, and "The Census Age Question: A Reply," by W. B. Bailey and J. H. Parmelee in the Quarterly Publications of the American Statistical Association, Nos. 90, 92, and 93 (June, 1910; December, 1910, and March, 1911).—TRANSLATOR.

[78] Thus, in statistics of industry and agriculture, groups of similar scope may not be formed, but rather, groups which correspond to certain types (small, medium-sized, and large industries, etc.). The magnitude classes of districts are often similarly demarcated.

cases the limits of the classes are ascertained through examination of the grouping of the items.

If magnitude classes are formed with previously established boundaries, the statistical series indicates the frequency or, in other words, how many items belong to the various classes; it will then be generally found, both in classes of equal and in those of unequal extent, that the number of items varies for the different classes. If, for instance, wage classes are formed: (1) up to $1.50, (2) $1.50 to $2, (3) $2 to $2.50, etc, the series will indicate the number of laborers in each class, and as a rule these numbers will differ from each other.

On the basis of a series which consists of magnitude classes with previously fixed boundaries, we may proceed to form " cumulative " magnitude classes, such as are frequently employed by English and American statisticians in dealing with wages. For this purpose the numbers of the items belonging to the several magnitude classes, starting from the lowest or the highest, are successively added, and the sums thus obtained are indicated. For example, instead of indicating how many laborers belong to each of the wage classes with fixed limits, we may, starting from the lowest, indicate how many laborers earn up to $1.50, how many up to $2, how many up to $2.50, etc.; or starting from the highest we may indicate how many earn $2.51 or more, $2.01 or more, $1.51 or more, etc. This latter method is the commoner. The sums signify in this case the numbers earning at or above a certain wage.[79-79a] Age data are often cumulated in the same way

[79] This kind of group formation was most freely employed in the Special Report of the Twelfth (U. S.) Census, entitled "Employees and Wages," prepared by Prof. Davis R. Dewey. Weekly wages are there presented in 50-cent groups, and the number of laborers, both in absolute figures and in percentages, is given for each class. The percentages are then worked over further according to the method in question, that is, proceeding from the highest wage class they are summed up successively. The resulting figures indicate what percent of

by indicating how many persons have passed the various age limits (so many persons at and above 0, 1, 2, 3, etc., years).

laborers receive certain wages or more. The column of these figures is called the "Cumulative Percentage Column."

This kind of group formation is especially valuable in comparing the wage conditions of different years. The "cumulative" percentages show clearly the evolution of wage conditions, while the absolute numbers and the percentages for the individual wage classes show apparently irregular shiftings from which no valid inference may be drawn. In the above-mentioned wage statistics of the American Census Bureau the wage conditions of the years 1900 and 1890 are represented side by side (p. xxvi). A few examples may be quoted. To the wage classes, \$8-8.49, \$8.50-8.99, \$9-9.49, \$9 50-9.99, \$10-10.49, there belonged in the years 1900 and 1890, respectively, the following percentages of laborers: 0.6 and 0.9, 0.1 and 0.3, 12.2 and 7.3, 2.9 and 1.1, 3.2 and 5.2. These figures are indecisive. The cumulative percentage column contains the following data for the corresponding "cumulative" classes: in 1900, 72.1% of laborers earned \$8 or more; in 1890, 78.1% (the percentage of laborers who earned less than \$8, therefore, increased from 1890 to 1900 by 6); in 1900, 71.5% of laborers earned \$8.50 or more; in 1890, 77.2%; in 1900, 71.4% of laborers earned \$9 or more; in 1890, 76.9%, etc. These cumulative percent figures leave no doubt that wages in 1890 were higher than in 1900.

[79a] There has been some difference of opinion in regard to the relative level of wages in the United States in 1900 as compared with 1890. Prof. H. L. Moore summarized the wage data of the census report on Employees and Wages in 30 selected industries, and found that the relative wages had declined from 100.0 to 99.6 in the decade. (See "The Variability of Wages" in the Pol. Sci. Quar., March, 1907.) The index number of the Bureau of Labor for rates of pay per hour shows an increase from 100.3 to 105.5 for the same period. Prof. W. C. Mitchell has examined the figures and methods used in the two computations (see "The Trustworthiness of the Bureau of Labor's Index Number of Wages" in the Quar. Jour. of Econs. for May, 1911), and has concluded that because the Bureau of Labor covered industries not included by Moore, and because the former took account of the reduction of working time by using hourly wages, "the results which Professor Moore deduced from Professor Dewey's report afford no reason for doubting that the Bureau of Labor's index number represents fairly the trend of wages in manufactuing industries."—TRANSLATOR.

Such series makes comparisons easy;[80] they have been much employed by Galton and other English statisticians for graphic representations. They have for this purpose the special advantage of producing a constantly rising curve, whereas the usual frequency curves rise and fall. The graphic representation of such series may also—as will be shown later—serve to determine the median and the mode.

If magnitude classes with previously fixed limits are to be formed, the average of the series may be used as a starting point, and the parts above and below the average formed into groups, in such a way that the average is the boundary between two of the groups. But this method is not often adopted; it is, however, frequently employed in the division of relative numbers which are to be represented in chart form; in such cases the parts above and below the average are generally colored differently and the varying intensity is indicated by shading.[81] If different phenomena are to be represented simultaneously in this way, the kind of group formation must, as a rule, be determined separately for each chart. The same color distinctions will then not signify the same degrees of intensity in the different charts. The same colors may indicate in one case great, in another case slight, deviations from the mean. In this way the reader, who looks at the different supplementary charts, easily gets false impressions. Cheysson in his *Album de Statistique agricole* tried to remedy this difficulty by dividing into groups, not the single relative numbers, but their distances from the mean, and to do this in the same way for a whole series of charts. Thus, similar colors meant equal distances from the mean. Ber-

[80] Cf. the note 79a.

[81] G. v. Mayr says (Theoretische Statistik, p. 112): "This manner of presentation must be pronounced wrong in principle. In resolving the total results of a district into geographical details there is no reason why such a decisive influence should be accorded to the magnitude of the general average."

tillon remarks in this connection: "This process is ingenious and very logical, but besides being laborious, it does not seem to have produced as good results as might have been expected, because it limits the means of expression, already very poor, which the shadings possess."[82]

From series of single observations there arises, as has been shown, by the formation of magnitude classes with previously fixed limits, a series which gives the number of items belonging to the individual classes. But often, simultaneously with this series, a second series is obtained by adding the magnitudes of the observation element, which are inserted in the several classes. Thus, if establishments are divided into magnitude classes according to the number of laborers, we obtain, first, a series which tells the number of establishments belonging to each class; but, secondly, we may also compute how many laborers in all are employed in the establishments of the different classes. If we divide farms according to their area, we learn how many farms belong to the various classes, and at the same time we may find out the total area in each magnitude class. In the same way, income statistics regularly give, not only the number of people investigated, but also the sum total of the incomes in the different classes.

But, as we have noted, other magnitude classes than those of previously determined extent may be formed from series of quantitative single observations. Instead of fixing the limits of the classes previously in order to ascertain the number of items in each class, the whole mass of items may be divided into equal parts and the limits may be thus determined. Each of these parts contains the same number of items, but the range in which these items lie, may vary. If a series is divided into four equal groups, the values forming the boundaries between the different groups are called quartiles; the quartile between the second and third group is at the same time the median of the series. If

[82] Cours élémentaire de Statistique, p. 142.

the series is divided into ten groups, the values which form the boundaries are called deciles; if there are a hundred groups, percentiles. In the last case the method of division is often called the "method of percentiles," or "Galton's method," after Galton, who has frequently employed it in anthropometry.[83]

In order to form equal parts, the items must be of equal weight; but, normally, relative numbers and averages possess varying weights, for which reason the method of percentiles is not applicable to series of the third group. It would be theoretically permissible with series of the second group, but would possess no significance worth mentioning, since such series, as a rule, do not consist of a sufficiently large number of members to make the application of the method profitable. Moreover, the series of the second and third groups, in order to be divided into percentiles, would have first to be arranged according to the numerical value of the items, whereas the natural order of these series is generally different. It follows, then, that the application of the method of percentiles is theoretically unobjectionable but at the same time of practical value only in series of quantitative single observations.

In order to form equal groups, it is, theoretically at least, necessary to know the whole series in detail. If the series already consists of magnitude classes with previously fixed limits and hence of unequal size, it is generally difficult to fix the position of the desired boundaries (quartiles, deciles,

[83] Cf. Galton, Natural Inheritance, p. 46, and his "Application of the Method of Percentiles to Mr. Yule's Data on the Distribution of Pauperism" (Journ. of the Roy. Stat. Soc., 1896, p. 392), also "Assigning Marks for Bodily Efficiency" (Report British Association, 1899, p. 475); Geissler, "Über die Vorteile der Berechnung nach perzentilen Graden" (Allg. statist. Archiv, II, 2); Prof. John Dewey, "Galton's Statistical Methods" (Quarterly Publications of the American Statistical Association, New Series, No. 8, 1888); L. Gulick, "The Value of Percentile Grades" (Quarterly Publications of the American Statistical Association, New Series, No. 21, 1893).

percentiles, etc.). But if we have the original statistical material, it is just as easy to divide it into groups of equal size as into groups with predetermined limits. To be sure, the formation of groups with predetermined limits, and especially the formation of groups of equidistant limits, will in the majority of cases be more significant, and accordingly this method is much oftener applied than the method of percentiles.

Magnitude classes formed according to the method of percentiles have no relation to an average except in one case. This case occurs where an even number of classes of the same size is formed, since the boundary between the two middle classes will then be identical with the median.

Since series of magnitude classes are very common, and especially since quantitative individual observations are very rarely published in full detail but generally only in magnitude classes of some kind, the statistician often has the task of determining the average of a series of such classes. Cases are rare in which the desired average of such a series is the boundary of a magnitude class and may thus be taken directly from the series. Generally the series does not express the average, which therefore has to be computed. Even if the original material were accessible in all its detail, the computation of an average from it would frequently be so laborious that the statistician prefers to deal with the magnitude classes. But the computation of an average from magnitude classes is open to one grave objection. The computation of an average presupposes, theoretically, the knowledge of the actual individual values of the series or, at least, of the values of certain portions of the series. In computing the arithmetic and geometric means, all the items of the series must be utilized. The former is obtained by adding all the items and dividing the sum by the number of items (n). The geometrical mean is found by multiplying all n items, and taking the nth root of the product. A knowledge of all the values

of the series is, indeed, not necessary in order to obtain the median or the mode. Generally we can easily ascertain in what magnitude class these two values are contained, but to determine them more exactly, it is necessary to consider the grouping of the individual values in the class concerned.

In order to compute averages from series which contain only magnitude classes, we must, accordingly, have recourse to auxiliary methods, especially to hypotheses about the distribution of the values in the different classes. These auxiliary methods, which differ according to the kind of average sought, will be taken up in the discussion of the various averages.

CHAPTER VI

NATURE AND PURPOSE OF AVERAGES

G. von Mayr describes the function of averages as follows: " The structure of a social mass, which has found expression in a series, can only be thoroughly understood by a careful study of all its members. But the more numerous the members and groups of members, the more impelling is the desire for concentrated information, that is to say, for a *single simple expression* which contains in itself the net result of the whole series. This is the purpose of averages."[84] He speaks of an average also as a " short expression of the phenomenon which levels all differences of the individual members of the series."[85] Bowley expresses himself in a similar way: " By the use of averages complex groups and large numbers are presented in a few significant words or figures,"[86] and " The object of a statistical estimate of a complex group is to present an outline, to enable the mind to comprehend with a single effort the significance of the whole."[87]

According to these citations, the essential nature of averages consists in describing series of divergent individual values by means of a simple comprehensive expression. The question now arises, for what concrete, methodological purposes we need such simple expressions and to what ends we can apply them, considering their nature. As we shall show presently, various purposes may be pursued in the computation of averages.

[84] Theoretische Statistik, p. 98.

[85] Ibid. p. 84.

[86] Elements of Statistics, 2nd ed., p. 107.

[87] Ibid. p. 7.

An average may be computed for its own sake, merely to obtain a comprehensive, characteristic expression for a series of divergent values. But it is often found as a means to another end, mainly for purposes of comparison, or in order to judge the individual values, or in order to measure the dispersion of series.

A. THE COMPUTATION OF AVERAGES FOR THEIR OWN SAKE

Every average characterizes a series in a definite way and also gives a measure of the complex of causes affecting the phenomenon in question. An average may be computed with the sole purpose of obtaining that information which the average from its nature is able to transmit. This information differs, however, according to the kind of average. The arithmetic mean, the median, the mode, etc., each give a different kind of information about the series from which they are obtained. The special characteristics of the different averages and the kind of information they give about a series will be discussed in the scond part of this book.

B. AVERAGES FOR PURPOSES OF COMPARISON

1. GENERAL REASONS FOR THE APPLICATION OF AVERAGES FOR PURPOSES OF COMPARISON

To compare series of individual values of any kind is often very difficult, especially if each series consists of numerous members and if a considerable number of series are to be compared. The comparison is made easier if the series to be compared are compressed into a few magnitude classes. But frequently that is not enough. In the place of the different series, their averages are then taken (arithmetic mean, or median, or mode, etc.); these may then be

compared, so to speak, at a glance. Let us suppose that a comparison is desired of the age conditions of those who marry in different countries or in different social groups. Even if the age tabulation of the masses to be compared is available in absolute numbers, a comparison of these series is naturally out of the question. And even if series of subordinate numbers occur, the comparison will hardly be possible. Accordingly, the attempt will be made first to compress the series into a few magnitude classes. If a survey is still impossible, the average age at marriage for the different lands or groups of population will be computed and these values compared with each other.

Averages are very often applied when series of quantitative individual observations, which refer to different times, countries, or groups of population, are to be compared; instead of a detailed tabulation of the ages of people in different times, or belonging to different groups of population, the average ages are very simply compared; the same process is followed if wages, incomes, heights, etc., are substituted for ages. It is extremely difficult, for instance, to compare several mortality tables, but it is easy to compare the average, probable or normal lengths of life computed from those tables.

In comparing averages it should always be borne in mind that the comparison extends only to those qualities of the series which the average is capable of reproducing, but that the other differences between the series cannot be determined by the comparison of these averages. If, for example, we compare the modes (that is, the relatively most frequent values) of two series of wage data, then this comparison is of course restricted to those qualities of the two series which can be expressed by the mode. Therefore, that series which possesses the higher mode may at the same time have the lower arithmetic mean, and *vice versa.* Also series which agree in regard to an average may at

the same time have a different dispersion or grouping of the items about the average.

Often, averages are computed for series which refer to long, successive periods of time, and compared in order to determine whether the phenomenon in question is subject to a distinct evolutionary tendency. This question can frequently not be answered from the irregular fluctuations of individual years. But if averages are compared for longer periods, the unessential fluctuations of the individual years are eliminated and the essential changes revealed.

The comparison of demographic numbers is usually equivalent to a comparison of averages. Relative numbers are commonly computed exclusively for purposes of comparison, since absolute numbers are not generally suited for this. If we wish, for instance, to compare the birth rate of two countries, the comparison of the absolute numbers of births in the two countries would lead to no conclusion. Countries of different sizes, of course, do not have the same numbers of births. Only by dividing the births in each case by the population, and thus eliminating the influence of populations of different sizes, do we obtain comparable values.

The comparison of averages generally consists in merely placing them side by side; in large numbers it is useful to compute the difference between the two figures in order to spare the reader this task. Often the two figures are brought into a percent relation. Thus, the average daily sick rate is usually given as a percentage of the average number liable to sickness. Finally, the difference between the two figures compared may be given as a percentage of the larger or smaller one.

2. MEAN INDEX NUMBERS

A unique example of the application of averages for purposes of comparison in time relations is furnished by

mean index numbers. These are computed on the basis of statistically determined changes of parts of a mass, in order to find out the change, not directly measurable, which affects the mass as a whole.

The best known application of this method is the computation of mean index numbers of prices in order to compare the level for different years. The prices of individual commodities are subject to special influences, depending on conditions of production and sale and, therefore, frequently show quite independent fluctuations, or, it may be, even opposite tendencies. But it may be significant to determine the net result of all these numerous individual movements and thus to ascertain the tendency of the general level of prices. For this purpose, the movements of prices of the individual commodities are given relatively to the prices of a standard year or to the average prices of a standard period by means of ratios or index numbers, and from these an average called the mean index number is computed for each year.[88]

A very extensive literature has grown up concerning index numbers and the problems arising from them. Both the British Association for the Advancement of Science and the International Statistical Institute have gone into the question thoroughly. The question at issue is what average (simple or weighted arithmetic mean, geometric mean, median, etc.) should be used in combining the index numbers for the individual commodities. If the weighted arithmetic mean is chosen, the question arises, How shall we determine the "weights" to be applied? We will return to these questions in the second part of the book

[88] Some authors limit themselves to adding the individual indices; for instance, The Economist and Kral. An opponent of mean index numbers in general is the Dutch economist, N. G. Pierson (see his "Further Considerations on Index Numbers" in The Economic Journal, 1896, p. 127 ff.; cf. also the reply of Edgeworth, "A Defence of Index Numbers," ibid. p. 132 ff.).

when we consider the various kinds of averages. Furthermore, we must determine what commodities we are going to select. *The Economist,* for instance, considers only 22, Sauerbeck 35 (with 45 index numbers because of the repetition of certain important ones), Sótbeer 114, Falkner 223, and the United States Bureau of Labor 258. For these commodities, in turn, various prices may be used (average prices or single quotations, wholesale or retail prices), which again may depend on various sources, such as trade journals, statements of merchants, etc. The standard year or the standard period has also to be chosen. The selection of commodities, of the method of averaging, of the source of prices, and of the standard period must be largely determined by the conditions surrounding particular investigations. However, we cannot investigate all of these conditions here.[89-89a]

Mean index numbers may be utilized to determine other general tendencies besides the movement of prices. Thus,

[89] Cf. the article "Preis" and (especially) the chapter "Statistische Bestimmung des Preisniveaus" by R. Zuckerkandl in the Handw. d. Staatsw. (with bibliography); also the article "Index Numbers" by Edgeworth in Palgrave's Dictionary of Political Economy and the article "Price-Levels" in Mulhall's Dictionary of Statistics; see also Tabellen zur Wahrungsstatistik of the Ministry of Finance (Vienna), 2nd ed., Pt. II, 4th number (Prices, Wages, Purchasing Power of Money), p. 919 ff. The special statistical questions are treated particularly by Bowley in Elements of Statistics, Chap. IX. Cf. also "Bericht über die Tätigkeit des statistischen Seminares der Universität Wien im Wintersemester, 1903-1904," review by J. Schumpeter of the method of index numbers (Stat. Monatsschrift, 1905, p. 191 ff.) and Report on Wholesale and Retail Prices (London, 1903), Appendix II.

[89a] A recent summary of the literature on index numbers is given in the Canadian Department of Labor report on Wholesale Prices in Canada, 1890-1909. Prof. Irving Fisher has given a thorough discussion of "The Best Index Numbers of Purchasing Power" in Chap. X and Appendix of "The Purchasing Power of Money," Macmillan, 1911.—TRANSLATOR.

Bowley has used them freely to determine the changes in the level of wages for large districts on the basis of known changes in the wages of certain localities, and also to determine changes in the wage level of certain industries from what is known as to the movement of wages of certain occupations within those industries. For instance, he has presented the evolution of the wages of agricultural laborers of different parts of Great Britain and Ireland by indices, and taking the average of the latter he has computed the index number for the whole United Kingdom.[90]

In the same way Bowley has also used indices to show the tendency of the wages of book printers in various cities and then by computing the weighted arithmetic mean has obtained the index for the whole United Kingdom.[91] Bowley has also computed index numbers of wages in 26 occupations connected with the building of machines and ships for 18 centers of industry. He has then combined these index numbers into mean index numbers, first, according to occupation for the whole kingdom, second, without distinction of occupation for the 18 centers of industry. Finally, from these mean index numbers he obtained grand mean indices by computing the simple and weighted arithmetic means. The grand mean indices show the tendency of wages for the laborers of the entire machine and ship building industry of the whole kingdom.[92] Bowley has, furthermore, obtained comprehensive mean index numbers for the movement of wages in the British building industry.[93] George H. Wood, his collaborator,

[90] The Statistics of Wages in the United Kingdom during the Last Hundred Years, Pt. IV, "Agricultural Wages" (Journ. of the Roy. Stat. Soc., Vol. LXII, 1899, especially p. 568 ff.).

[91] Ibid. pp. 708-715.

[92] Ibid. Vol. LXIX, 1906, p. 158 ff.

[93] Ibid. Vol. LXIV, 1901, p. 106 ff.

has also computed averages from indices of wages of miscellaneous occupations.[94-94a]

Wood has also undertaken to express the development of the consumption of the English population by means of general index numbers.[95] The consumption of the population and any change in it can, of course, only be obtained from observations regarding the consumption of single articles. Wood has, accordingly, represented the consumption per head of the population of England, 1860-1896, of the more important articles (flour, meat, rice, coffee, sugar, tea, tobacco, etc.) by means of index figures, fixing the average consumption of his standard period, 1870-1879, at

[94] "Changes in Average Wages in New South Wales, 1823-1898," Journ. of the Roy. Stat. Soc., Vol. LXIV, 1901, p 327 ff. (based on the data in T. A. Coghlan's Wealth and Progress in New South Wales). Prof. Falkner observes ("Die Lohnstatistik in der Theorie und in der Praxis," Allg. Stat. Archiv, Vol. VI, 1st half-vol., 1902) that he applied the method of index numbers to wages even before Bowley in the Aldrich Report of 1893, and that Sir Robert Griffen recommended special indices for wages in his review of index numbers made for the International Statistical Institute, 1887. Isolated index numbers for wages are also to be found in Salaires et Durée de travail dans l'Industrie française of the year 1897 (for example, Vol. IV, p. 277). Recently total index numbers have been employed in the wage statistics published annually by the American Bureau of Labor in such a way as to embrace the indices for the movement of wages in the various occupations by industries and finally for the total industry. (See Bulletin of the Bureau of Labor, Washington, July, 1907, p. 22 f.). Cf. also the suggestions of W. C. Mitchell regarding the application of index numbers in wage statistics in Publications of the Amer. Stat. Assoc, Vol. IX, p. 325 ff.

[94a] The English Board of Trade has recently used the method of index numbers for comparison of wages and hours of labor, rents and housing conditions, retail prices of food and the expenditure of working-class families on food for the United Kingdom (see Report Cd. 3864), with similar items for Germany (Cd. 4032), France (Cd. 4512), Belgium (Cd. 5065), and the United States (Cd. 5609).—TRANSLATOR.

[95] "Some Statistics of Working Class Progress since 1860," Journ. of the Roy. Stat. Soc., Vol. LXII, 1899, p. 639 ff., especially p. 654 f.

100. From these index numbers Wood computed the arithmetic mean and five different weighted means, in which the changes in the total consumption of the population are expressed.

Similarly, Wood has also measured the changes in the amount of aid given to the unemployed of the various English trade unions for the years 1860-1896 by means of index numbers (taking as his standard period 1882-1891), and has combined these index numbers into an average.[96]

Neumann-Spallart has attempted to solve a much more comprehensive problem by the use of mean index numbers, namely, that of finding a "measure of the variations in the economic and social condition of nations." He distinguished for this purpose four groups of symptoms, the *first two* being economic, as affecting production and trade, the *third,* socio-economic, affecting consumption, emigration, banks, etc., the *fourth,* moral, affecting birth rate, percentage of illegitimate births, suicides, crimes, etc. His plan, as presented at the meeting of the International Statistical Institute in Rome (1887), was to compute index numbers for the different symptoms and then to combine these into averages, first, for the several groups of symptoms, second, for the single countries and, finally, for the six countries, England, France, Germany, Austria, Belgium, and the United States, taken together.[97]

This plan, whose accomplishment was unfortunately prevented by Neumann-Spallart's death, arouses grave misgivings. The indices from which he wished to compute averages do not refer to complementary parts of a homogeneous totality (as, for instance, the prices of the various commodities which make up a general level of prices), but to very different social phenomena; each of these phenomena may be symptomatic of certain economic and social

[96] "Trade Union Expenditure on Unemployed Benefits since 1860," Roy. Stat. Soc. Journ., Vol. LXIII, 1900, p. 88 ff.

[97] Cf. Bulletin de l'Inst. intern. de Stat., Vol. II, Pt. I, pp. 150-159.

conditions, but it is very questionable whether, by taking all these phenomena together, we obtain a valid measure for the general social condition and its changes.[97a]

3. MEANING OF THE POSTULATE OF THE GREATEST POSSIBLE HOMOGENEITY FOR THE COMPARISON OF AVERAGES AND RELATIVE NUMBERS

The postulate, previously stated,[98] of the greatest possible homogeneity of series or masses lying at the basis of averages and relative numbers, has special significance in the comparison of averages and of such relative numbers as are, in reality, averages. The comparison of only those averages and relative numbers which refer to homogeneous series or masses yields reliable conclusions; the comparison of other averages and relative numbers may easily mislead and is only reliable under special conditions. This will be demonstrated in the following pages, for the most part with the same examples as have already been employed[99] to establish this postulate in general in connection with the nature of averages and relative numbers.

[97a] Roger W. Babson has recently combined a miscellaneous lot of statistics into a single series of index numbers which he calls "Summary Barometer Figures" He combines twenty-five series of statistics, which are grouped under the twelve headings. building and real estate, bank clearings, business failures, labor conditions, money conditions, foreign trade, gold movements, commodity prices, investment market, crops and commodity statistics, railroad earnings, and social conditions. Cf. Business Barometers for Forecasting Conditions, published by Roger W. Babson, Wellesley Hills, Mass. Jas. H. Brookmire of St. Louis, in the publication entitled The Brookmire Economic Charts, publishes similar figures. However, Brookmire weights his individual indices according to the importance that they are supposed to have in indicating economic crises. (See "Methods of Business Forecasting Based on Fundamental Statistics" by J. H. Brookmire in the Am. Econ. Rev., March, 1913.)—TRANSLATOR.

[98] P. 65 ff.

[99] P. 65 ff.

(a) *Postulate of the greatest possible homogeneity in the comparison of averages obtained from series of individual observations*

Series of quantitative single observations which are not homogeneous are sometimes so because they contain cases which do not show the element of observation at all, the causal complex being inoperative in such cases. The size of the arithmetic mean of such a non-homogeneous series depends on the ratio of the items having a positive to those having a zero measurement. The same thing is true, *mutatis mutandis,* of certain other averages as, for instance, the median.[100] A comparison of averages from such heterogeneous series is evidently unreliable. That one of the series in which the element investigated has the greater values may yield the smaller average, simply because the series contains relatively more items with zero measurements (null items). Let us take as an example the average marital fecundity. If we compare the average number of children for two countries from all the marriages, fruitful and unfruitful, it may happen that the country in which the marriages, when not entirely unfruitful, produce the greater number of children, may show the smaller general marital fecundity because of the larger percentage of unfruitful marriages.

The elimination of null items is, of course, only feasible where individual observations occur. For example, in computing the average consumption of meat or alcohol for the total population, it is impossible to eliminate those persons or classes who do not take part in the consumption. Great difficulties must evidently arise from this fact in making comparisons. If, for instance, two countries or cities are compared, it may happen that, in consequence of different percentages of those not taking part in the consumption, the persons actually concerned in the country

[100] See above, p. 66.

or city with the smaller average may consume more than in the country or city with the larger average. The percentage of individuals not concerned may evidently vary in different places, since it depends on many demographic conditions, not only on the sex and age structure of the population, but also on various social and psychological factors. A trustworthy comparison of general averages will only be possible when we may assume that the ration of the individuals concerned to those not concerned is, on the whole, the same in the masses to be compared. This prerequisite will hardly ever be fulfilled in geographical comparisons, but, on the other hand, fairly often in time comparisons which do not extend over too long a period. Thus, we may properly draw certain conclusions from the figures for the average consumption of alcohol or meat for a country or city for a term of years.

Series of quantitative single observations may also be regarded as non-homogeneous, if among the items, so far as they are positive, special parts may be distinguished which are governed by different and independent causes. The size of the average (both arithmetic mean and certain other averages) obtained from such a series depends essentially on the proportion in which the various more homogeneous parts stand to one another.[101] Hence, the comparison of such averages may easily be misleading. Let us consider the comparison of wages in two districts by means of general average wages. Suppose that all laborers without distinction of sex have been included, although sex in the districts in question has an influence on the amount of wages. Let us assume that in district A 50% men and 50% women are employed, the wages of the men average $15, those of the women $5, the average for all laborers being $10; in district B the women constitute only 10% of the total number of laborers, the wages of the men average $12, those of the women $4, so that both

[101] See above, p. 70.

men and women have considerably smaller wages than in district A. Nevertheless, the general average wage in B is $11.20 because of the smaller percentage of women, while in A it is only $10. Whoever compares merely the general averages will be tempted to assume that in B the wage conditions are better, whereas exactly the opposite is the case. As a matter of fact, false conclusions often occur in cases which correspond to the scheme just illustrated.[102]

This scheme may also easily be extended to the case where within the masses to be compared there are more than two heterogeneous parts. For example, within a mass of laborers several categories may occur with various wages. Thus, according to Austrian statistics, the mine laborers consist of pickmen and their helpers, other adult miners, adult day laborers, boys and women laborers; the wages of these categories vary greatly, decreasing in the order mentioned. Now if we compute average wages for all the miners for different years, the comparison of these averages might readily be misleading. Though the wages of each category had remained the same from year to year, yet the average wages of all the miners might have fallen in case the more poorly paid categories had increased. The wages of each category might indeed have increased and yet the general average have fallen.[103] A comparison of

[102] The average age at marriage in Germany offers an example which corresponds to this scheme but which substitutes a time for a space comparison. This average has fallen in the last few years. Industrial laborers and the agricultural population form a large proportion of those who marry, the former with a very low and the latter with a relatively high age at marriage. Because of the industrial development of Germany the former constitute an ever increasing quota of the total population. The average age of those marrying must necessarily fall merely by reason of this circumstance, even though neither the industrial laborers nor the agricultural population actually marry earlier than formerly. (Cf. G. v. Mayr, Bevölkerungsstatistik, p. 401 f.)

[103] The following would be a similar case: The average number

averages of series in which there are heterogeneous parts is only allowable, therefore, on the assumption that the heterogeneous constituents are present in equal proportion in the masses to be compared. This condition will most often be fulfilled in the case of time comparisons which do not extend over too great a period.

From the above examples it may be seen how easily erroneous conclusions may be drawn from statistical comparisons. Even the trained statistician is exposed to errors, when he does not realize that the averages which he compares refer to non-homogeneous masses. If the heterogeneous groups which compose a mass cannot be separated from each other (for instance, because the criterion, according to which the items were to be distinguished, was not considered when the observation was made), and if it is also not certain that these groups are represented in like proportion in the masses, then even the best statistician is unable to arrive at a result. This also explains how the same statistical material, when differently handled and grouped, seems frequently to prove exactly opposite assertions. Hence a certain popular skepticism about statistics "which can prove anything." The statistical method really demands unusual accuracy and conscientiousness. Not seldom does the trained statistician recognize the insufficiency of his material and, refraining from positive assertions, confines himself to merely hypothetical conclusions, while the untrained layman draws false conclusions from the statistical data.

of inhabitants per house may have increased from one census to another. May we conclude from this that a house, on the average, now contains more people and that the population is living closer together? Certainly not, for larger houses may have been built between the two censuses. Similarly, it would be insufficient to compare the average number of pupils per school at two different times without reference to possible changes in the number of classes of which the schools consist.

(b) *Postulate of the greatest possible homogeneity in the comparison of averages from series whose members indicate the size of masses limited in a definite way*

Among the series of the second group, whose members indicate the size of definitely limited masses, time series are the most important. From time series, as we have already mentioned, averages for parts may be computed and compared with each other in order to ascertain whether the series shows a definite tendency. It may, under certain conditions, be possible to form and compare parts of time series which are homogeneous in themselves but differ from each other in regard to their causation. In such comparisons everything depends on the kind of division. If the line of demarcation is wrongly drawn, it may be that no conclusion is possible, or else a false conclusion may be reached. If, on the other hand, we succeed in comparing homogeneous time divisions, valuable information may be obtained.

(c) *Postulate of the greatest possible homogeneity in the comparison of relative numbers*

The same difficulties as in the comparison of averages from non-homogeneous series recur in the comparison of relative numbers which have arisen from the subdivision or interrelation of non-homogeneous masses.

The comparability of relative numbers may, first of all, be impaired by the fact that in computing them the parts with zero measurement were not eliminated.[104] The numerical value of the relative numbers will then depend on the frequently varying proportion of the null items. For instance—to cite first a subordinate number—the proportion of the unmarried in a country depends largely on the number of those who have not yet reached the marriage-

[104] See above, p. 73.

able age. If, therefore, we compare the percentage of the unmarried in two countries, the higher percentage in one country may be due to a relatively larger number of children, while among those capable of marrying the unmarried may not be more numerous. A trustworthy conclusion cannot be drawn until the percentage of the unmarried is determined and compared for those of marriageable age in the two countries.

For similar reasons the general frequency figures, which arise from the interrelation of definite events (births, marriages, crimes, etc.) with the total population, are hardly comparable. Their numerical value depends chiefly on the ratio of the positive to the null parts, which ratio may vary from case to case. In the country A the marriageable population may show a higher marriage rate than in the country B. Yet the country A may have the lower general marriage rate, if the unmarriageable age classes are relatively larger than in B. Thus, the mere comparison of such general frequency figures may be wholly misleading, and for purposes of comparison those "specific" frequency figures are regularly to be preferred which arise from the interrelation of the events with the parts of the population having a positive measurement.[104a] General frequency figures are comparable only when the percentage of the null parts of the population is the same in the cases compared. This is hardly ever the case in comparing different countries, and is permissible in time comparisons for the same country only for limited periods.[105]

[104a] There are, however, certain purposes for which the null items need not be eliminated. The crude rates are the very figures we desire when, for instance, we wish to compare results for two centuries regardless of the various causes contributing to those results. Then, too, in many cases statisticians are forced to use crude rates because the data for the computation of more refined rates are not available.—TRANSLATOR.

[105] Cf. Westergaard's discussion of cases in which the general mar-

In computing relative numbers not only is the elimination of null parts desirable but also the division of the masses into more homogeneous parts, primarily, as already mentioned,[106] in order to obtain relative numbers corresponding to a definite, unified causal complex, and also, as will now be shown, to make comparisons easy.[106a]

The numerical value of relative numbers based on nonhomogeneous masses depends essentially on the proportion in which the various parts of different subdivisions or of different intensity stand to one another. If this proportion differs in the masses to be compared, then a comparison is very difficult if not quite impossible. We shall give two examples, one for subordinate and one for coordinate numbers.

Let us assume that the comparison of two populations (exclusive of those not yet of marriageable age) in regard to conjugal condition has indicated that population A possesses a considerably larger percentage of widows than population B. The question now arises whether a definite conclusion is permissible on the basis of this comparison. To ascertain this, let us try to form more homogeneous parts in the two masses to be compared. Since conjugal condition varies, as experience shows, with age, let us divide the two populations according to age and compare the cor-

riage figure is applicable for time comparisons, in Die Grundzüge der Theorie der Statistik, pp. 149 and 161.

[106] See above, p. 75 f.

[106a] See the admirable paper by G. Udny Yule for illustration of the method of correcting crude birth rates to obtain birth rates that take into account both the number of wives in the population and their ages ("The Changes in the Marriage- and Birth-Rates in England and Wales during the Past Half Century; with an Inquiry as to Their Probable Causes," Jour. of the Roy. Stat. Soc., March, 1907). Also see Newsholme and Stevenson on "The Decline of Human Fertility in the United Kingdom and Other Countries as Shown by Corrected Birth-Rates" (Jour. of the Roy. Stat. Soc., March, 1907).—Translator.

responding age classes; it may result from this that in each single age class the population B has more widows than A, though A as a whole has the larger quota. This apparent contradiction may easily arise from a different age division. The older periods of life, of course, show in both populations many more widows than the younger. The older age periods of A accordingly have many more widows than the younger age periods of B. If now the older age periods in A are relatively more populous than in B, there necessarily results for A, when all age classes are united, a larger percentage of widows than for B, although B in each individual age class has more widows than A.

In order to make clear the difficulty of the comparison of coordinate numbers for non-homogeneous masses, we may refer to the best known case, that of the general death rate. Let us, for the sake of simplicity, consider merely the difference of mortality according to age. If we try to compare the mortality of two countries by means of the general death rate, it may happen that country A has the higher general rate than country B, although in B the individual age classes in themselves may show the higher mortality. This apparently contradictory condition would occur if, for instance, in A those age classes, which (like the early years of infancy) naturally have a greater mortality, were relatively stronger than in B.[106b] Thus, we may not without further investigation attribute a higher general death rate to less favorable mortality conditions. To make the general death rates of different countries comparable is the purpose of the method of mortality index, which has given rise to much discussion.[107]

[106b] Westergaard gives an illustration of the same apparently contradictory condition as that cited above in his Mortalität und Morbilität; the general death rate of clergymen is greater than that of railway employees; but the death rates of clergymen by age groups are much lower than the corresponding rates of railway employees.—TRANSLATOR.

[107] Cf. below, p. 160 f.

4. INVESTIGATION OF CAUSATION BY COMPARISON OF AVERAGES AND RELATIVE NUMBERS

The comparison of averages and relative numbers possesses an especial importance as a method of investigating causes, where from the difference in numerical value of two or more averages or relative numbers we infer a difference of causation in the phenomena thus characterized.

If we compare averages or relative numbers which refer to different concrete geographical districts or time periods, and if we establish a considerable difference in numerical value, we perceive from this divergency that influences, either quite different or of varying strength, have affected the masses compared; but we have no further information about the difference which exists between the causes operating upon the two masses, and we do not learn to what influence to attribute the difference of the averages or relative numbers. If, for example, we ascertain that in country B the average length of life is shorter, and the death rate higher than in A, or if we find that in the same country the average length of life has increased and the death rate fallen from decade to decade, this proves indeed that there exists between the masses compared a difference in regard to their causation, but we cannot infer what the nature of this difference is. We must try to answer this question by further statistical methods or in some non-statistical way.

The case is different, however, when two masses are compared which are distinguished from each other by a definite qualitative or quantitative criterion (for instance, sex, occupation, age, etc.), or by an abstract space criterion (altitude, temperature, soil, etc.), or by an abstract time criterion (season, etc.). If in such a case a considerable difference is established between the averages or relative numbers computed for the masses in question, it is permissible, under definite conditions to be discussed later,

to attribute this difference to the difference of criterion in those masses. Thus, if men and women, or different occupations, or age classes show a varying mortality or different length of life, we may, under definite conditions, conclude that sex, or occupation, or age influences mortality or length of life.

This is the method which corresponds in statistics to that inductive method which J. S. Mill in his *Logic* called the "method of difference." If we compare two groups of phenomena which differ from each other in one respect both in antecedent and consequent, we conclude, according to this method, that the members in which the compared sequences differ stand to one another in causal relation. If one group consist of antecedents A B C D and of consequents a b c d, and a second group of antecedents B C D and consequents b c d, it follows that A is the cause of a. Statistics has a similar problem in the comparison of averages and relative numbers for statistical masses which differ in regard to a definite objective factor. Thus, from the fact that two statistical masses differ, on the one hand, in regard to the sex of the individuals in question and, on the other hand, in regard to mortality, we conclude that sex influences mortality. But there are essential differences between statistical material and the material of the natural sciences, to which the inductive method is regularly applied. From a conclusion in natural science a general law is obtained which admits of no exception. Such a conclusion becomes invalid if a single case is known which contradicts it. If we have observed that several individuals of the same species possess a definite mark or character, and if we infer from the given evidence that the whole species possesses this mark, this conclusion is invalidated by a single instance to the contrary. It is quite different in regard to statistical conclusions. These never hold good for single cases, but only for statistical masses (aggregates). The proposition, that the average length of

life of men is shorter than that of women, by no means asserts that all men die younger than women, but only that on the whole, on the average, women reach a greater age. This proposition is wholly compatible with an extremely large number of opposite cases. Therefore, this method of causal investigation by comparison of averages and relative numbers cannot be regarded as the usual inductive process—as the majority of statisticians seem to regard it—even though it is very closely related to this process. We must, rather, agree with A. A. Tschuprow, who has demonstrated in his very instructive article, "Die Aufgaben der Theorie der Statistik,"[108] that the statistical method should be given an equal place with the inductive process in the system of formal logic. Both methods, according to Tschuprow, have the same purpose—to establish "natural laws" by refining the raw material of observation; but they are applied under different conditions; the inductive methods serve to disclose an invariable connection between cause and effect; the statistical method, on the other hand, is the essence of such methods of investigation as render possible the study of the looser causal connections characterized by the plurality of causes and effects.[108a]

In order to infer a difference in cause from the divergency of the numerical values of two averages or relative

[108] Jahrbuch für Gesetzgebung, Verwaltung und Volkswirtschaft im Deutschen Reiche. Edited by Gustav Schmoller. Vol. XXIX, Pt. II, 1905, p. 27.

[108a] For discussions of the scope and methods of economics and statistics see Venn's Logic of Chance, Ernst G. F. Gryzanovski's paper "On Collective Phenomena and the Scientific Value of Statistical Data" (published as No. 3, Vol. VII, Third Series, of the Publications of the Am. Econ. Assoc., August, 1906), H. L. Moore's article on "The Statistical Complement of Pure Economics" in the Quar. Jour. Econs., November, 1908, also his Laws of Wages (Macmillan, 1911), Keynes' Scope and Method of Political Economy, and the article on "Method of Political Economy," in Palgrave's Dict. Pol. Econ.—TRANSLATOR.

numbers, the divergency must be considerable and important. Minor differences will naturally not justify such an inference. The statistician will have to judge in the individual case if a difference is significant. For the elementary mathematical statistician the decision is more or less a question of subjective estimate. The calculus of probability makes it possible for the mathematical statistician, on the other hand, to determine in many cases with what probability the difference between the two values compared may be regarded as accidental, or what probability there is for the existence of a different causation.

The number of objective criteria, according to which statistical masses may be differentiated, is known to be extremely large, and in very many cases the masses, which differ in regard to a definite criterion, produce in fact averages and relative numbers of different numerical values. For instance, the death rate is different for each of the sexes, for various age classes, occupations, etc.; it varies with the season and, apparently, with the altitude. Similarly, the average length of life, the average age at marriage, the marital fecundity of different groups of the population differ from one another. Various phenomena of moral and economic statistics also show characteristic differences for different groups of population. The determination of such divergencies is one of the most important tasks of statistics, and statisticians must continually endeavor to disclose new characteristic differences by an ever increasing differentiation of statistical material. Often in the statistical investigation of causes a causal connection is presumed on the basis of some extraneous knowledge. This connection is then to be statistically proved. In this case an hypothesis is first set up and then verified by a tentative division of the statistical material (by an "experimental formation of groups") and by the comparison of parts which are distinguished from each other in regard to the factor which is considered causal.

But the statistical method is only able to prove causality within definite limits. The comparison of averages or relative numbers may show that definitely characterized masses, which differ from each other in certain respects, yield relative numbers or averages of different numerical values. The result of the comparison is a coincidence of definite facts which, indeed, allows us to infer a causal connection, but which gives no information as to which fact is the cause and which the effect. This question must be decided in the individual case on the basis of further knowledge, perhaps by means of further statistical investigations.[108b]

The decision will generally be very easy. Thus, if we

[108b] Concerning this point Prof. H. L. Moore holds that "Economic events are not arrayed in linear connection, the one event following the other in direct series, as was frequently assumed by the classical economists. It was an idle controversy that Malthus and Ricardo conducted upon the question whether the abundance of food increases the population or the multitude of consumers increases the supply of food. Social phenomena are interrelated, are mutually dependent, and the appropriate method of treating such a form of interdependence is the use of a system of simultaneous equations in which the equations are equal in number to the unknown quantities in the problem" (Laws of Wages, p. 2). However, there is a controversy going on at the present time that appears to hinge on just this question of the order of economic phenomena. It is the question of the relation between the quantity of money and prices. J. L. Laughlin is the leader of the school holding that variations in prices precede and *cause* variations in the amount of currency. He says, "When the price is fixed, the credit medium by which the commodity is passed from seller to buyer comes easily and naturally into existence and, of course, for a sum exactly equaling the price agreed upon multiplied by the number of units of goods. . . . That is, the quantity of the actual media of exchange thus brought into use is a result and not a cause of the price-making process" (Bul. Am. Econ. Assoc., April, 1911, pp. 28, 29). The classical theory is that variations in the quantity of the media of exchange precede and *cause* variations in the price-level. Irving Fisher is the chief modern protagonist of this theory. He holds that the price-level "is not cause but effect" (Bul. Am. Econ. Assoc., April, 1911, p. 38).—TRANSLATOR.

find that the more wealthy classes of the population show a lower mortality, we may evidently designate the wealth as cause and the lower mortality as effect. The case will not be so simple if we determine, for instance, a coincidence of greater wealth with fewer children. Is the wealth the cause of the small number of children, as is often asserted, or is it not also conceivable that many families have attained to wealth by reason of the smaller number of children? Another interesting example is the coincidence of greater infant mortality with higher birth rate. It seemed self-evident that, in general, the reason was that as the number of children per family increased the care and attention bestowed on each must necessarily decrease. But the very opposite causal connection has been proved, at least for a certain group of cases. Geissler has ascertained[109] the period elasping between the births of two successive children in the same family for 26,429 families of Saxon miners, and he has differentiated these measurements according to whether the firstborn child died or remained alive. He has demonstrated that the interval was shorter if the older child died. Evidently, if a child died, the desire was aroused for another child, or else the checks, which usually delayed the begetting of another child, vanished. This fact shows that not only the number of children may influence mortality but also, under certain circumstances, that mortality may influence the number of children. Here is evidently a case of mutual influence, of interdependence, such as the social organism so often shows. In many other cases there is no immediate causal connection at all between the two facts whose coincidence is statistically shown; they are both under the influence of a deeper common cause. Thus, if we find that criminals are on the average smaller than non-criminals, a direct connection between bodily size and criminality is, naturally, excluded, but deeper lying common causes may exist to

109 Zeitschrift des. königl. sächs. statistischen Bureaus 1885, p. 24.

which physical inferiority and moral depravity may be ultimately attributed.

When we speak in statistics of the investigation of causes, we are, of course, never concerned with investigating the causes of single cases or individual events. The statistical investigation of causes can refer only to characteristic conditions, in which masses of individuals (or other units) are found. These conditions are often, so to speak, merely the frame in which there are operative individual causes of various kinds and intensities, about which the statistical comparison itself yields no information. As regards these individual causes, other sciences, according to the object of the investigation, may supply information; so far as they affect a considerable number of individuals they may be determined by means of further detailed statistical investigations; but they may also remain quite unknown. Thus, the comparison of the mortality of boys and girls or of legitimate and illegitimate children develops the undoubted fact that there is a greater mortality among the boys and among the illegitimate children; popularly speaking, therefore, membership in the male sex and illegitimacy are designated as the causes of the greater mortality. The real immediate causes which result in the individual deaths of boys more than of girls and of illegitimate more than of legitimate children—are not disclosed by the statistical comparison; they can only be discovered by medical research or by further statistical investigation of details. The same thing is true if we follow the influence of wealth upon various demographic phenomena, for instance, on mortality. To ascertain that poverty increases mortality does not disclose the causes immediately operative. But by further and more special investigations we may inquire in what way wealth affects the causes of death, whether the same diseases take a different course with the wealthy and the poor, etc.

The situation is the same in numerous other cases; for

example, if we measure the influence of conjugal condition, occupation, the seasons, etc., on various demographic, moral-statistical, and economic phenomena. Even though in all these cases the direct individual causes are not discovered by the statistical method in question and even though we may not speak of the determination of real, exact, causal laws, yet by this method regularly recurring differences are discovered. These differences found to exist between definitely characterized groups of individuals (or other units) cannot be disclosed by other methods than the statistical, and their discovery forms the starting point of further and more thorough investigations and, consequently, may also furnish a basis for measures of social reform.[110-111]

If the difference between two averages or relative numbers is to be causally related to a precise difference in the masses compared, then these masses must be assumed to differ from each other only in that precise way, but to agree in all other respects. The conclusion that there is a causal connection is only permissible on the assumption that other things are equal. If this assumption is not true, if the masses diverge also in regard to another criterion besides the one used to differentiate them, then we

[110] Interesting examples are the investigations as to the influence of the various occupations on mortality and on diseases, and the attempts to get at the actual causes such as stooping posture, dust, fumes, dampness, etc.

[111] While there are plausible explanations for most statistically determined differences (such as the influence of economic position, of the seasons, etc.) the fact of the different sex-ratio in living births and still-births, in legitimate and illegitimate children, seems quite inexplicable to the layman. Lexis (Abhandlungen zur Theorie der Bevölkerungs- und Moralstatistik, VII, "Das Geschlechtsverhältnis der Geborenen und die Wahrscheinlichkeitsrechnung," p 166 ff.) has tried to relate these differences to different percentages of early births; G. v. Mayr (Bevölkerungsstatistik, p. 188) has connected the greater excess of boys in the country as compared with the city with the relatively greater inbreeding and has explained in the same way the excess of boys among the Jews.

cannot decide what difference in the masses is the cause of the divergence of the averages or relative numbers.

If we find, for instance, that the male and female laborers of a definite district have different average wages, but if we know at the same time that the men are in different occupations than the women, we cannot decide whether the difference in the average wages is to be attributed to difference of sex or to difference in the categories of work. If we compare the mortality in various occupations and know that those belonging to these occupations have a different age grouping, we are not justified in attributing the difference in mortality to the occupation, since it may also come from the different age grouping.

The prerequisite that other things be equal is seldom completely fulfilled; a certainty that it is fulfilled can never be attained. The conclusion of causal connection will, therefore, always be merely hypothetical and more or less probable.[112-112a]

[112] v. Inama-Sternegg ("Neue Beiträge zur Methodenlehre der Statistik" in Staatswissenschaftliche Abhandlungen, 1903) distinguishes the progressive method (experiment), by which we proceed from cause to effect, and the regressive method (observation), by which we infer the cause from the effect. He asserts that while the direct proof of a causal connection is obtained by the progressive method the regressive method leads only to an explanation that is a mere hypothesis.

Although the great majority of statisticians recognize in the investigation of causality one of the most important, though most difficult tasks of their science, theoretical opponents have not been wanting. Napoleone Colajanni (Statistica teorica, p. 265) mentions Bodio as such. G. Staehr ("Einige Bemerkungen über die statistische Methode," Bulletin de l'Institut international de Statistique, Vol. IV, No. 1, Pt. II, p. 288 ff.) denies that statistics is competent to determine causes; he thinks it is an empirical-descriptive but not an inductive-analytical science.

[112a] Prof. H. L. Moore holds that the argument of statistics is purely utilitarian or pragmatic in character. He quotes from Jevons's Theory of Political Economy as follows: "The deductive science of economics must be verified and rendered useful by the

The assumption that other things are equal will possess the greatest probability of being true in comparing highly homogeneous masses, for instance, if we compare the average wages of male laborers, on the one hand, and female laborers, on the other hand, of the same occupation and of a single category of work, or the mortality of individuals of the same sex and age of two different occupations. But, as already emphasized, complete homogeneity is never reached. On the contrary, statistics is regularly concerned with masses which cannot be called homogeneous. Two non-homogeneous masses differ solely with respect to the criterion by which they are differentiated (the condition of a conclusion) only in case they are made up *in a like manner* of groups of the more homogeneous constituents, the groups corresponding to the other criteria coming into consideration; only in such a case can the comparison of non-homogeneous masses lead to the isolation and measurement of a definite influence. For instance, if we compare the average wages of all the male and female laborers of a definite district, a conclusion as to the causal influence of sex upon wages will only be possible if it is certain that the male and female laborers are similarly distributed in the different occupations and categories of work, and that they possess the same age classification, etc.; from the comparison of the mortality of those belonging to different occupations, the influence of occupation will only be manifest if the sex and age classification, etc., of the persons compared is the same. The investigation of causality by a comparison of statistical averages and relative numbers is thus limited by prerequisites which must be strictly examined and which, unfortunately, are often not fulfilled. But statistics has to depend on the material at its disposal and cannot like physics, for instance, devise experiments

purely empirical science of statistics." (See "The Statistical Complement of Pure Economics," Quarterly Journal of Economics, November, 1908, p. 16.)—TRANSLATOR.

in order to observe and measure the effect of a factor which may be added or eliminated at pleasure. The statistical investigation of causes, therefore, gives rise to errors only too frequently, and a perfect theoretical certainty can never be obtained.

The conscientious statistician must, accordingly, renounce entirely any conclusion as to causality if the masses compared by him differ not only in regard to the one criterion employed but also exhibit other differences whose influence cannot be accounted for numerically or be ignored as trifling. Influences impossible to estimate occur, unfortunately, only too often. Thus, a proof of the influence of wealth upon mortality is generally impossible because the members of the different economic classes belong to various occupations. We encounter a similar difficulty if we try to determine the influence of various religions on the morality of the population, since religious differences generally coincide with national differences.

Mention should also be made of the controversy which was once waged in regard to the influence of conjugal condition on mortality. From the fact of the lower mortality of married people, several authors (among them Bertillon) concluded that the influence of marriage was beneficial. Block opposed this view, pointing to the factor of selection and showing that many people do not marry on account of physical ailments and that those who do marry are accordingly healthier than those who do not marry; therefore, lower mortality for the married should be expected *a priori.* A strict statistical proof would only be obtained if, on the one hand, married, and on the other hand, unmarried people of equal physical condition (and of the same occupations, wealth, etc.) could be compared in regard to their mortality,—a thing which is impossible.

The factor of selection is important in many other fields, as, for example, in the comparison of the demographic conditions of various occupations. The choice of an occu-

pation is often made by a kind of natural selection, since certain occupations require definite physical conditions. Hence, we cannot conclude from the different mortality of those belonging to the occupations that the occupation is the cause of the difference.

In the above cases we have dealt with the comparison of masses diverging in regard to a quantitative or qualitative criterion which must have been fixed when the statistics were obtained. For instance, if the wages of men and women are compared, the sex of the individual laborers must have been indicated when the wage data were obtained, so that these wage data could be divided subsequently, into two masses according to the sex of the laborers. The case is somewhat different, when, in order to establish a causality, averages of time or place series (or parts of them) are compared, which series at the same time differ from each other in a quantitative or qualitative respect. The series to be compared are, in this case, not differentiated according to a factor considered when the data were obtained, but according to either a non-statistical criterion or one taken from some other statistical data. Thus, we may compare the years before and after a definite fact, for instance, the promulgation of a new law, in order to see whether this fact had an influence on a definite phenomenon and, if so, of what importance it was.[112b] Or, we may compare periods of time or districts which differ in regard to their economic condition, to which evidently some causal significance belongs; for example, we may compare periods of economic prosperity and depression in regard to unemployment, mortality, criminality, etc.

The considerations indicated above for the comparison of masses differentiated in a merely quantitative or quali-

[112b] A fine illustration of such a comparison may be found in the article by G. H. Wood on "Factory Legislation Considered with Reference to the Wages, etc., of the Operatives Protected Thereby" (Jour. Roy. Stat. Soc., Vol. LXV, pp. 284-324).—TRANSLATOR.

tative way are also true for the comparison of time and place masses with simultaneous qualitative or quantitative differences; for instance, the consideration that the statistical comparison establishes the existence of a causal connection but gives no certainty as to which fact is cause and which is effect, holds true here. Furthermore, time and place comparisons of the kind in question can only lead, like purely quantitative and qualitative comparisons, to a conclusion of causality on the assumption that other things are equal, and in the concrete case we must inquire whether this assumption is permissible.

C. AVERAGES AS STANDARDS FOR JUDGING ITEMS

The items of a statistical series are often comparea with its average in order to ascertain whether particular items are above or below the average and how far they diverge from it. This determination may be of great significance for the judgment of the items, since great deviations from the average indicate, as a rule, the existence of special causes. The information, which may be obtained by the comparisons of items with the average, varies, however, according to the kind of series involved.

If in a series of quantitative single observations an item differing greatly from the average is found (for instance, the wages or the length of life of a certain individual), it is certain that this difference is to be attributed to special causes, but the statistical method in question is unable to reveal the nature of these special causes. If we have a series of the second group, whose members indicate the size of definitely limited masses, or those series of the third group which consist of relative numbers or averages referring to different time or place masses, it is also impossible to determine a definite causal connection by comparing an item with the average. Thus, if a definite mem-

ber of a time series consisting of absolute or relative numbers or averages shows a striking divergence from the arithmetic mean, this will indeed indicate that special causes have affected the mass at the time of the item in question. But these special causes can only be discovered by means of other statistical investigations or in a non-statistical way.

The case is different when an item of a qualitative or quantitative series of the third group is compared with the average of the series. The item refers, in such a case, to a part which differs from the totality in regard to a definite qualitative or quantitative criterion. If, in such a comparison, a considerable difference appears between the relative number (or average) referring to the part and the relative number (or average) characterizing the totality, then we may infer—other things being equal—that the factor which especially characterizes the part is the cause of the difference in the values compared. For example, if we find that those of a certain occupation have a mortality considerably above the average, we attribute this divergence from the average—other things being equal—to that occupation; if we find that a disease appears more frequently in a certain age class than on the average for the whole population, we ascribe—other things being equal—to the age in question an influence on the frequency of the disease. This is simply a variety of the investigation of causality discussed in the preceding chapter by comparing averages and relative numbers; this variety is marked by the fact that the values compared are not coordinate—as, for instance, the death rate of men as compared with that of women—but are in the mutual relation of item and average. The principles established in the preceding chapter may therefore be applied, with the necessary changes, to the comparison of the relative numbers and averages here involved.

The comparison of a specially characterized part with

the totality is often made in order to ascertain whether a certain factor exerts an influence in one way or another. But this comparison does not give directly the precise extent of this influence, for the following reason: The totality includes also the specially characterized part; the particular influence to be measured has, therefore, affected the numerical value for the totality and hence does not find precise expression in the comparison. The smaller the part in relation to the totality, the less important will be the disturbance. But if the part is a large percent of the totality, the value for the totality is already considerably influenced and the comparison becomes of little or no significance. Thus, the average height of criminals is often compared with the average height of the total population. In order to measure exactly the connection of criminality with height, we should really have to compare not criminals and the total population but criminals and non-criminals. If criminals, as it appears, are on the average smaller than the total population, the non-criminals must on the average be taller than the total population; accordingly the difference between criminals and non-criminals must be greater than the difference between criminals and the total population. But since criminals are only a very small fraction of the total population, the average height of the total population is only imperceptibly influenced by them, and the comparison of the part (criminals) with the totality (whole population) is sufficient. It would be quite different if, for instance, we were concerned with the influence of sex on height. No one would think of comparing the average height of persons of one sex with the average height of the total population; as a matter of course, the two sexes would be compared directly with each other.

In practical statistics there are, however, some cases where a statistical item is compared with an average of the same logical content. In the computation of the aver-

age in such a case, the item in question is not taken into account. For example, the Austrian harvest statistics gives the crop yield of the individual years as a percentage of the average of the preceding ten years (thus, the crop of 1904 in relation to the average of the period 1894-1903). Among the motives which may lead to such a proceeding there is perhaps the consideration mentioned above, that the comparison of an item with an average, whose size has been affected by the item itself, does not clearly express the strength of the particular causes influencing that item. A similar case often occurs when the level of prices of various years is compared by means of total index numbers. A single year is frequently not chosen as a basis of comparison, but the average of a definite number of years, the years of the "standard period."[113] With this average are compared not only the years belonging to the standard period—in which items are compared with their average—but also the years preceding or following the standard period.

D. THE FUNCTION OF AVERAGES IN THE MEASUREMENT OF THE DISPERSION OF SERIES

Averages may, as we have already explained, serve as standards for the judgment of items. If not merely a single item of a series is judged by the average, but rather all the items of the series, we obtain a picture of the grouping or dispersion of the whole series about the average. To obtain such a picture is very often necessary for the judgment of statistical series. The dispersion of a series of individual observations indicates the degree of the variability of an individual character (for instance, height, wages, etc.), whose measurement is very often of the great-

[113] Sauerbeck takes the years 1867-1877 as the standard period, The Economist the years 1845-1850, Soetbeer the years 1847-1850, Conrad both the years 1879-1883 and the years 1879-1889, etc.

est significance; the dispersion of a time series gives a measure of the constancy or variability of a phenomenon in the course of time; the dispersion of a geographical series reveals the variations to which a phenomenon is subject in the districts under consideration, etc.

Hence, averages are often computed to serve as a point of departure for the investigation of the grouping of the items. This investigation is of prime importance since statistical series show the greatest multiplicity in regard to the distribution of their items, and scarcely two series exist of similar and equally great dispersion. This investigation also makes it possible to divide the statistical series into different sub-classes. Many statistical series show a more or less regular dispersion. These are the series which can be particularly well expressed by averages. The averages from such series may be supplemented by special values, which mark the dispersion of the series about their averages. Those series may be best designated in this way whose members are symmetrically grouped about the average according to the law of chance. The indication of the average and of a measure of dispersion (such as the average deviation or the standard deviation) are enough in such a case to characterize the series in its entirety. Other series are not, indeed, distributed symmetrically about their averages, but yet the grouping of the items about the average may be brought under an extended law of chance. Series whose members show no sort of regular grouping about their means may, moreover, be divided into more homogeneous parts possessing a regular dispersion. Among the series of non-symmetrical dispersion about the average those deserve a special interest which show a characteristic regular conformation in some other way.

In the above we have dealt with the case where an average is computed for the purpose of serving as a starting point for the measurement and judgment of the dispersion of the series; in such a case the measurement of the dis-

persion is the real object, the computation of the average simply a means to this end. But the determination of the dispersion is also very useful where an average is used for an independent, primary purpose, such as the comprehensive characterization of a series, or for purposes of comparison. An average in itself in no way expresses the dispersion of the series from which it arises, and yet the methodological value of each average, especially the question whether it may be regarded as a " typical " average or not, depends on this dispersion. Therefore, when averages are employed, data should also be obtained in regard to the dispersion of the series.

In measuring the dispersion of statistical series of individual observations, we may start from various averages, the arithmetic mean, the mode, etc. The description of the various methods of measuring dispersion can, therefore, only follow the discussion of the various kinds of averages.

PART II

THE VARIOUS KINDS OF AVERAGES

CHAPTER I

SYNOPSIS

Each average, as has been said, serves to characterize a series by a single numerical expression. This characterization can be accomplished in various ways and consequently various kinds of averages are possible. A complete enumeration of all the numerical values which may characterize a series is, of course, impossible.[1]

Only those kinds of averages will be considered in the following which are actually used in statistics. Such are the arithmetic mean, including the weighted mean, the geometric mean, the median, and the mode. The peculiar properties of each of these means will be investigated, that is, what they indicate about the series in question, how they are computed, in what departments of applied statistics they are useful, and the like.

[1] Fechner says ("Über den Ausgangswert der kleinsten Abweichungssumme," Abhandlungen der kgl. sächsischen Gesellschaft der Wissenschaften, Vol. XVIII, p. 74): "By a mean of given items we understand a value which can be derived from these items according to definite principles and which falls among the values of the items, or more shortly, a value which is a function of the items falling between the minimum and maximum items. Consequently, there are an indefinite number of means as there are an unlimited number of definite principles or functions of the kind described. Only those means deserve special mention which have special mathematical or empirical interest." Messedaglia also remarks ("Calcul des valeurs moyennes," Annales de démographie internationale, 1880, p. 388) that one can conceive of an unlimited number of means of various kinds. Messedaglia mentions that the Roman philosopher and mathematician, Boëtius (470-525 A.D.), enumerated ten means in his work, De Arithmetica.

Aside from the means named there are others which originate from the items of the series through the assumption of other mathematical principles, for example, the harmonic mean,[2] the contraharmonic mean,[3] and the mean square or quadratic mean.[4] In addition, Fechner men-

[2] Messedaglia developed especially the harmonic mean and its relations to the arithmetic and geometric means. ("Calcul des valeurs moyennes," Annales de démographie internationale, 1880.) The formula for the harmonic mean (M) between the two values, a and b, is $\frac{2\,a\,b}{a+b}$. This formula does not hold for more than two values. For the case of several values, $a_1, a_2, a_3 \ldots . a_n$, Messedaglia gave several formulas of which the following is the best known:

$$M = \frac{n}{\frac{1}{a_1} + \frac{1}{a_2} + \frac{1}{a_3} + \ldots\ldots + \frac{1}{a_n}}.$$

(Cf. Messedaglia, p. 397, and Blaschke, Vorlesungen über mathematische Statistik, p. 71.) The harmonic mean of several values is always less than the arithmetic or geometric mean of these values. Thus the values 1 and 2 have the harmonic mean 1.33, the geometric mean 1.41, and the arithmetic mean 1.50. Of the three means of the same set of values the geometric is always the geometric mean of the other two. Given any two of the three means the third may be found from them. (Messedaglia, p. 390.)

[3] The formula for the contraharmonic mean is

$$M = \frac{a_1^2 + a_2^2 + \ldots . + a_n^2}{a_1 + a_2 + \ldots\ldots + a_n}.$$

The contraharmonic mean of several values is always greater than the harmonic, the arithmetic, and the geometric means. The contraharmonic mean of 1 and 2 is 1.66. The arithmetic mean of any set of values always equals the arithmetic mean of the harmonic and contraharmonic means of the same set. Any one of these means can, therefore, be computed from the other two. (Messedaglia, p. 393 f.)

[4] The quadratic mean of the values $a_1, a_2, a_3 \ldots . a_n$ is the square root of the arithmetic mean of the squares of the values $\left(\sqrt{\frac{a_1^2 + a_2^2 \ldots . + a_n^2}{n}}\right)$. (Cf. von Bortkiewicz, Das Gesetz der kleinen Zahlen, p. 9.) This mean is used by Gauss in the theory of error in computing the mean error. The quadratic mean

tions the " Scheidewert," the " schwersten Wert," and the " Abweichungsschwerwert."[5] Although none of these means, except the standard deviation, are actually used to characterize statistical series, yet some, for instance the harmonic and contraharmonic means, have been investigated and discussed because of their interesting mathematical properties.

Of the means used in statistics the arithmetic average is the most important. Mathematicians and statisticians have always used it and investigated its properties. The geometric mean is, likewise, familiar to mathematicians but seldom used in statistics.[5a]

The median and mode have been developed recently, mainly by Fechner and the English statisticians and widely applied by them. Yet, Messedaglia, undoubtedly one of the foremost statisticians of his time, in his paper entitled "*Calcul des valeurs moyennes,*" published in the *Annales de démographie internationale* in 1880, did not mention these at all. The three " classical " means which were his principal objects of investigation were the arithmetic, geometric, and harmonic means, all of which, Messedaglia stated, were known to Plato and Aristotle and considered by Boëtius, the Roman mathematician and philosopher, in his *De arithmetica.*

of several values is always greater than the arithmetic mean and less than the contraharmonic mean; it is identical with the geometric mean of the two last named means. (Messedaglia, pp. 394, 402.)

[5] See Fechner, Kollektivmasslehre, pp. 160, 172-181.

[5a] The geometric mean was applied to price statistics by W. Stanley Jevons in his well-known study, A Serious Fall in the Value of Gold Ascertained and Its Social Effects Set Forth, published in 1863 (reprinted by the Macmillan Co. in 1884 in Investigations in Currency and Finance). F. Y. Edgeworth (Jour. Roy. Stat. Soc, Vol. XLVI, p. 714), Francis Galton (Proc Roy. Soc., Vol. XXIX, p. 365), Donald McAlister (Proc. Roy. Soc., Vol. XXIX, p. 367), and A. W. Flux (Quar. Jour. Econs., Vol. XXI, p 613) have discussed the use of the geometric mean in statistics.—TRANSLATOR.

A statistical average falls into one of two groups according as all, or only a portion, of the items of the series are required in its computation. On the one hand, the computation of arithmetic and geometric means depends upon *all* the items of the series. On the other hand, the mode and median are single definite items chosen to characterize the series because of their positions in it.

Fechner has offered another classification of means which will be given although it is of more significance to mathematics than to statistics. His groups are, first, power means, second, means "which contest with power means for their name," and third, combination means. A power mean is a value, the absolute sum [5b] of like powers of the deviations of the items from which is a minimum.[6] The best known power means are the median and the arithmetic average for which the absolute sum of the first powers and squares, respectively, of the deviations of the items are minima. The second group of means are those of the form $M_n = \sqrt[n]{\frac{\Sigma a^n}{m}}$.[7] In this group the arithmetic mean is of the first order, and the mean square of the second order. In the third group of "combination means" the arithmetic mean is likewise of the first order, the geometric mean being of the fourth order.[8] According to Fechner it is to be noted that "the arithmetic mean is common to

[5b] "Absolute sum" means that all deviations are to be considered positive.—TRANSLATOR.

[6] "Über den Ausgangswert der kleinsten Abweichungssumme," Abhandlungen der königl. sachs. Gesellschaft der Wissenschaften, Vol. XVIII, p. 37 f.

[7] Ibid. p. 74. The notation in the formula has the following significance:

n = power
Σ = "sum of such terms as"
a = item of the series
m = number of items —TRANSLATOR.

[8] Ibid. p. 75 f.

these three groups defined by distinct principles and therefore, if we do not have special reasons for using another mean, this fact gives us a reason for choosing it as the most satisfactory mean."

Fechner also has undertaken to obtain averages (which he calls main values) on the basis of logarithmic treatment of series of single observations.[9] From the logarithms of the various members of a series he computed averages (so-called logarithmic main values) in order to investigate in this way "the logarithmic deviations" of the items, i. e., the deviations of the logarithms of the items from the "logarithmic main values." The "logarithmic main values" which he computed, were first, the mode of the logarithms of the items (which must not be mistaken for the logarithm of the mode computed from the items themselves), second, the median of these logarithms (the "logarithmic median"), and third, their arithmetic mean. From the logarithmic main values Fechner secured the natural values corresponding to these logarithms from the logarithm table. He called the natural or numerical value of the logarithmic mode "the proportional mode" because it is the characteristic of this value that in equal proportional distance from it in both directions more values are united than in the same proportional distance from any other value. This "proportional mode" differs from the arithmetic mode. Fechner discovered further that the numerical value belonging to the logarithmic median coincides with the median obtained directly from the items themselves. The natural value of the arithmetic mean of the logarithms of the items is identical with the geometric mean of the items.

Unfortunately the terminology in the field of averages is, as yet, uncertain. In German the words "Mittelwerte"

[9] Cf. Kollektivmasslehre, pp. 24 f., 79-83, 339-351; see also "Über den Ausgangswert der kleinsten Abweichungssumme," p. 14 f.

and "Durchschnittswert" may each indicate merely the arithmetic average, or averages in general. In English the existence of two expressions, i. e., "average" and "mean," has led to attempts to make a distinction between these two expressions.[10] Bowley thinks that it would be best to use "average" for a purely arithmetic concept, such as the average duration of life of a mixed population. This average duration of life does not hold for any constituent homogeneous group of the population and is only a short expression for the result of a certain arithmetic operation; the word "mean," however, ought according to Bowley to be used for objective quantities such as the mean height of the English people around which mean all the different measurements group themselves with definite regularity. Bowley's proposition apparently is a result of the distinction between "atypical," and "typical" means; the former would be "averages," the latter "means." The consequence of Bowley's terminology, however, is that there would be no English word left for the general idea of the mean. In fact the English idiom is very uncertain. The expressions "average" and "mean" are used generically as well as to indicate the arithmetic mean in particular.[11] In this treatise the words "average" and

[10] Elements of Statistics, 2nd ed., p. 107.

[11] Thus, Venn in his paper "On the Nature and Uses of Averages" (Jour. of the Royal Stat. Soc., 1891, p. 430) uses the word "mean" for the arithmetic mean; on the other hand the word "average" is very often employed for this mean. Thus in the special report of the United States Bureau of the Census on Employees and Wages (1903, p. xxvii), it is used in contrast to the median. The theoretical English statisticians frequently use the words "average" and "mean" in a generic sense and they select more specific terms for the different types of means.

Likewise the Italian terminology appears to vary. For instance, Colajanni, to be sure in opposition to the great majority of Italian statisticians, includes only the arithmetic and geometric means under the term "Valori medî," and calls the median and mode "other

" mean " are both used to denote the general concept unless the context clearly indicates reference to some particular mean.

values" which may be used to characterize series (Manuale di Statistica teorica, p. 181).

The French statisticians usually employ "moyenne" for average in the broader sense and more specific terms for the various kinds of averages (moyenne arithmétique, géométrique, etc.); yet the arithmetic mean is many times simply designated as "moyenne."

CHAPTER II

THE ARITHMETIC MEAN

A. THE SIMPLE ARITHMETIC MEAN, OR, SHORTLY, ARITHMETIC MEAN

1. CONCEPT AND QUALITIES OF THE ARITHMETIC MEAN

The simple arithmetic mean, the most widely known and used statistical mean, is computed by dividing the sum of the items by their number. The arithmetic mean denotes the size which the items would have if, the sum total remaining unchanged, they would all be made equally large. The statement of this value carries with it important information about the series from which the arithmetic mean was computed.

From the manner of computing the arithmetic mean it follows directly that the sum of the positive deviations from the arithmetic mean is equal to the sum of the negative deviations from that mean. According to another mathematical theorem the arithmetic mean of a series of items is characterized by the fact that the sum of the squares of the deviations of the items from the mean is a minimum.

From the definition of the arithmetic mean it follows that its value will be affected by a change in any member of the series. This is not the case with other means. The median and the mode, for instance, may remain unchanged even if considerable parts of the series are changed, since these means are not computed from all the items but are found by choosing one item to represent the series because of the characteristic position of that item in the series.

Therefore, median and mode can only be found in series the members of which are arranged according to magnitude. The arithmetic mean, however, does not presuppose any definite arrangement of the items and the same value is obtained by adding the items in any order.

Furthermore, the arithmetic mean has this advantage over the median and the mode that it can be computed from every series of items while the latter can only be obtained in series of individual observations, and even in these cases the mode cannot be computed unless there be a decided point of concentration.

2. DISTINCTION BETWEEN STATISTICAL SERIES WITH REFERENCE TO THE COMPUTATION OF THE ARITHMETIC MEAN

The consideration of the various conditions under which arithmetic means are computed, leads us back to our division of statistical series into three groups. These were, first, series of individual observations; second, series the members of which indicate the size of quantities that are limited in a certain way (constituents of a totality); third, series the members of which characterize definitely limited quantities (parts of a larger whole) in a certain manner by relative numbers or means.

From the series of the first two groups arithmetic means can be computed directly by dividing the sum of the items by their number. The mean can be computed directly from these series even if they consist of subordinate numbers. In a series of the *first group* subordinate numbers indicate what *percent* of the single cases fall into the various numerical classes. Then the average magnitude of the element under observation is computed by treating the subordinate numbers like absolute numbers and by dividing the sum total of the series by 100. If wage data are under consideration and if 20% of the workmen whose wages were ascertained receive $20.00

per week, 30% receive \$22.00, 20% receive \$24.00, 20% receive \$26.00 and 10% receive \$28.00, then, in order to obtain the average wage, the various wage items are multiplied with the corresponding percentage and the sum is then divided by 100. In this way it will be found that the average wage is \$23.40.[12]

If a series of the *second group* consists of subordinate numbers, they indicate what percent of the totality of a higher order falls to the constituents. In order to find the percentage of the totality which on the average falls to a constituent—this percentage depending merely on the number of the constituents—it is merely necessary to divide 100 by the number of constituents. If we have a series of subordinate numbers which indicate what per cent of all the deaths of a year occur in each month, then the average percentage per month (needed to ascertain what months are above and what are below the average) is found by dividing 100 by 12, giving 8 3%.

We now come to series of the *third group,* the members of which are relative (subordinate or coordinate) numbers, or averages. However, we shall not here consider these series if their members are other than relative members or arithmetic means (for instance, medians or modes) since only higher means of the same kind (i. e., also medians or modes) may be contrasted to these items, while the computation of an arithmetic mean from the items is excluded.[13] But in no case should a simple arithmetic average be computed directly from the items of a series of the third group. The members of such series as a rule refer to different quantities (constituents) and, consequently, are of different weight, while the relative importance of the different members is not clear from the series itself. If we have a series of death rates for different years, terri-

[12] $(20 \times \$20) + (30 \times \$22) + (20 \times \$24) + (20 \times \$26) + (10 \times \$28) = \$2{,}340 = 100 \times \$23.40$.

[13] P. 17 f. and p. 22 f.

tories, professions or ages, then different "weights" must be given to these numbers because they correspond to fractions with different denominators on account of the change in the population in the course of years, or the various population of the territories, or various numbers in the professions or ages. If we treat the single members of such a series as equal and compute a simple arithmetic mean directly we arrive at a wrong result. In such a case the mean must under no condition be computed directly from the items, but independently, on the basis of the corresponding data for the entire quantity in question. Thus with reference to the death rates for the various parts of a country or groups of population the mean death rate for the entire population must be computed by bringing the number of the whole population into independent relation with all the deaths having occurred. The average death rate per year is obtained by dividing the total number of deaths for the period in question by the sum total of the yearly population, or by dividing the average death rate by the average population for the period. The value thus computed is the weighted arithmetic mean of the members of the series, that is, the death rate for the whole population is the weighted arithmetic mean of the death rates for the single provinces, or the different groups of population to which the items refer. The constituents have contributed to the resulting average according to their weights. Thus the general average wage of the workmen of a certain territory forms the weighted arithmetic mean of the average wages for certain categories of workmen, and the general average duration of life forms the weighted arithmetic mean of the values for the average mean duration of life of the people belonging to different groups of the population. Consequently, if we desire to compute the true arithmetic mean for a totality from its constituents (not having data complete enough to use the method described above) it is necessary to estimate the

relative importance of the constituents and find their weighted average.

Although it is the rule that relative numbers or averages which refer to differences of time, space, and quality or quantity (constituents of a larger totality) are of unequal weight, yet cases may occur where these inequalities are so trifling that they need not be taken into account. Especially in time series in the field of population statistics the items very often differ only slightly in weight, if the population has not changed considerably in the course of the years under consideration. In such cases a simple arithmetic mean may be computed directly from the items, if necessary, without resulting in a large error. If the items are of entirely equal weight, then the value computed independently for the totality is the simple arithmetic mean of the items and is identical with the value which is obtained directly from the items. In such cases the method of computation is merely a question of convenience, dependent upon the material at hand.

3. COMPUTATION OF THE ARITHMETIC MEAN

The computation of the arithmetic mean, from its known items, is purely mechanical, the simple arithmetic operations required being known to everybody. However, the statistician is quite frequently confronted by the task of computing averages from series [14] that do not exhibit the original items individually but that consist of classes, i. e., the series merely indicating how many items there are between certain limits. It cannot be seen from such series in what manner the items belonging to the single classes are distributed between their limits. The computation of the arithmetic mean, however, presupposes, at least theoretically, the knowledge of all the single members of a series, since they must be added. In order to be able to

[14] Cf. p. 89 f.

compute arithmetic means from series that consist of classes we usually resort to hypotheses as to the grouping of the items within the classes, and then we use the values corresponding to that particular hypothesis in the computation. The hypothesis that the items are distributed uniformly in the single classes is most frequently used in this connection. Therefore the mid-value of each class is taken as the average of all the items belonging to that class, and is used in the computation of the arithmetic mean for the whole series.[15]

The actual grouping of the items within the different classes, of course, never agrees completely with the hypothesis of uniform distribution. If classes of wide limits are given, the hypothesis of uniform distribution of the items, in most cases, is incorrect. An example of the difficulties that must be overcome in such a series is given in the computation of the average age of marriage from the data as published by most statistical bureaus.[16] In these publications the ages of those marrying are usually presented in classes of several years each. Since the frequency of marriage changes considerably with age, it is clear that those marrying cannot be distributed uniformly within the given age classes. If classes of ten years each are formed the hypothesis of the uniform distribution cannot be used even for one single class without resulting in a considerable error. In the class that contains the ages 20 and less than 30 years, the males will probably be more densely crowded together in the years at the end

[15] Now and then other hypotheses are used in the computation of the arithmetic mean of a series consisting of classes. Thus in the special report of the U. S. Bureau of the Census Employees and Wages (1903, p. xxvii), arithmetic means are found in the computation of which "the lowest wage in each wage group was taken as the exact wage for each individual in the group" (ibid. note 1). This is a simplified procedure, but theoretically not quite correct.

[16] Compare with this the remark of G. v. Mayr in Bevölkerungsstatistik, p. 402.

of the period than in the years at the beginning, while the majority of the females will probably belong in the first half of this period. In the class 30 and less than 40 years of age, both sexes undoubtedly will be more densely crowded in the first part and the farther we progress the rarer they will be. Therefore, it is incorrect to assume that all those in the class 20 and less than 30 years are 25 years old on the average and all those 30 and less than 40 years, 35 years old on the average. In fact the average age of males of the first class (20 and less than 30 years) is higher, and that of females of this class as well as the average age of both sexes in the second class (30 and less than 40 years) is lower than results from the hypothesis of uniform distribution. In order to compute the average age of all those marrying we must obtain, first of all, the average age in each class. But how may these class averages be estimated without the aid of more detailed data?

Theoretically, of course, the difficulty in computing the average of a series which consists of classes, is always the same, no matter if these classes are narrow or wide. But the errors that may result if the hypothesis of uniform distribution is used for wide classes are far greater than for narrow classes, for instance, age classes of one year. The age distribution of the living is usually given in one-year classes. But even here the hypothesis of uniform distribution in the classes is not always free from objections. In the higher age classes the distribution within the single year is certainly not uniform but decreases towards the end and, consequently, the hypothesis of uniform distribution would result in an average age which is somewhat too high. However, this error will be comparatively small.[17] Therefore it often serves the purpose to first divide larger classes by interpolation into smaller

[17] Compare with this G. v. Mayr, Bevölkerungsstatistik, p. 84.

classes and then to compute the arithmetic mean from the latter by using the hypothesis of uniform distribution.

Among the series consisting of classes there are such whose first and last classes are limited only in one direction. With such series the computation of the average is especially difficult. The following represents such a series for the ages of those marrying: under 20 years, 20 and less than 25 years, 25 and less than 30 years, 30 and less than 40 years, 40 and less than 50 years, 50 years and more. The lower limit of the first class and the upper limit of the last class are unknown. However, the items which belong to the first class (under 20 years) cannot go below the legal age of marriage, while the items of the class " 50 years and more " are limited by the maximum duration of life. But these are extreme limits and the items undoubtedly do not extend quite so far in reality. The items " under 20 years " will all be close to 20 years and the items " 50 years and more " close to 50 years. Other details are not known. Therefore it is necessary to estimate rather arbitrarily the average age of those marrying " under 20 years " and " 50 years and more " as a preliminary to the computation of the average age of *all* those marrying.

An interesting example of the computation of an average from a series consisting of classes that has no maximum limit, is given in the treatment of the statistics of tourists in the publication of the Austrian Treasury Department *Daten zur Zahlungsbilanz.*[18] The length of the sojourn of tourists in certain places is registered as follows: up to 3 days, from 3 to 7 days, from 1 to 2 weeks, from 2 to 3 weeks, from 3 to 4 weeks, from 4 to 5 weeks, from 5 to 6 weeks, longer than 6 weeks. In the publication quoted the hypothesis of uniform distribution is used for all classes with the exception of the lowest (up to 3 days) and of the last class without maximum limit (longer than 6 weeks),

[18] Tabellen zur Währungsstatistik, 2nd ed., Pt. II, No. 3, p. 829.

i. e., in all classes, with the exception of the two named, it is assumed that the average length of the sojourn may be expressed by the arithmetic mean of the two limits of the classes in question. In the class "up to 3 days," which, under the above assumption, would give the arithmetic mean of "2 days," an average length of sojourn of 1.2 days is assumed. The average sojourn of people who at registration were put in the class "longer than 6 weeks" is supposed to be 50 days. On the basis of these averages assumed for the single classes (and under the assumption of an average sojourn of 2 days for those persons about whose length of stay no declaration was made) the total average for the entire series, i. e., the average length of sojourn of all tourists together, is found to be 8.5 days.[19]

Complete statistical series, that is those whose items are given in detail, as well as series consisting of classes, are sometimes subjected to adjustment, in order to remove the more or less accidental unevenness in the formation of the series or in the form of the curve resulting from graphic representation of the series. For statistical series usually exhibit irregularities in the details, even if a certain characteristic formation can be recognized. These can be traced back to the inevitable accidental errors which every empirical determination of a value shows, to the limitation of the field of observation and to the imperfections of the observation (for instance, incorrect declaration of age). In addition to these there may occur special disturbances in the normal course of the observed phenomena (for instance, epidemics in the case of mortality statistics).[20]

[19] Fechner has developed a special mathematical procedure for the computation of the arithmetic mean of a series without superior and inferior limits (Kollektivmasslehre, § 128 ["Supplementarverfahren"]). This procedure, however, is applicable only under the supposition that the series corresponds to the asymmetrical Gaussian law.

[20] Von Bortkiewicz in the article "Ausgleichung der Sterblichkeitstafeln" in Handw. d. Staatsw.; compare also Czuber, Wahr-

The adjustment may be done graphically by constructing a curve which supposedly represents the essential characteristic traits of the structure of the series. There are also various mechanical methods of adjustment [21] as well as those based on mathematical functions. The latter methods are used especially in the adjustment of mortality tables.[22-22a]

Although by adjustment of a series we intend in the first place to improve its general formation, yet this adjustment has a special importance for the determination of the means from the series in question. By the adjustment the items of the series are modified and it may hap-

scheinlichkeitsrechnung, No. 196, "Ausgleichung von Tafeln," p. 392 f.

[21] The mechanical methods of adjustment or graduation often depend upon computations of averages. Wittstein's method is as follows: He takes the arithmetic mean of each 5 successive items throughout the whole series to be adjusted and puts it in the place of the middle item of the group. Woolhouse and Karup proceed by finding five values for every item of the series to be adjusted, one of these values results from the observation itself while the other four originate from interpolation. The arithmetic mean of these five values is taken to be the adjusted value (cf. Czuber, Wahrscheinlichkeitsrechnung, p. 403 ff.).

[22] Of the voluminous mathematical literature on the methods of adjustment may be mentioned especially: Blaschke, Die Methoden der Ausgleichung von Massenerscheinungen, Vienna, 1893, and the same, Vorlesungen über math. Statistik, Leipsic, 1906 (particularly Pt. VI); cf. also Czuber, Die Wahrscheinlichkeitsrechnung, No. 196 "Adjustment of Tables," No. 198 "Mechanical Methods of Adjustment" and "Graphical Adjustment"; Bowley, Elements of Statistics, 2nd ed., pp. 254-258; Westergaard, Die Grundzüge der Theorie der Statistik, pp. 130-136, and Die Lehre von der Mortalität und Morbilität, pp. 111 f. and 202 f.

[22a] Allyn A. Young gives a bibliography on methods of adjusting age data in "The Adjustment of Census Age Returns" (Western Reserve Bulletin, November, 1902). The same writer also gives a brief discussion and bibliography of this subject in Bulletin 13 of the Bureau of the Census (pp. 47-53). Newsholme describes easy graphic methods of adjustment in his Vital Statistics.—TRANSLATOR.

pen that the arithmetic mean computed from the adjusted series is not exactly the same as that which would result from the non-adjusted series. Large differences, however, must not be expected, since the uneven points which were removed by the adjustment probably would have counterbalanced each other in the computation of the arithmetic mean. At any rate, the size of the arithmetic mean of an adjusted series may depend in a certain measure on the manner of the adjustment and on the method chosen.

When computing the arithmetic mean from a series, as mentioned, the items of the series are added and then divided by the number of members. Therefore, in order to be able to compute the arithmetic mean from a series, either all the single members of the series, the sum of which is to be found, must be known, or if this is not the case, estimated values must be substituted for them. The knowledge or the estimation of the single members, however, becomes unnecessary, if their sum and number be given. In such a case it is sufficient to divide the former value by the latter in order to compute the arithmetic mean. To be sure, the isolated mean which we find in this way gives only limited information about the series. A thorough insight is possible only when the items are ascertained and arranged in a statistical series from which we may compute the arithmetic mean, as well as other means and the dispersion of the items around the mean.

4. APPLICATION OF THE ARITHMETIC MEAN

As is well known, arithmetic means are used frequently in all departments of statistics. First of all the arithmetic means used in the field of population statistics must be mentioned: the mean duration of life (of the new-born or of people at certain ages), the average age of the living, the dead, and those marrying, the average number of children per family. Various fundamental questions concern-

ing the manner of computation and the usefulness of these averages have already been touched upon in previous chapters. However, it would lead us too far afield to investigate the scientific importance of these averages and the conclusions which can be drawn from them with reference to the peculiarities of population statistics.

As an illustration of the use of the arithmetic mean in other fields we may mention: the average height of people, which plays an important rôle in anthropological statistics, the average temperature, and the average barometric height in the field of meteorology, the average number of people living per domicile, the average size of landed properties or of agricultural establishments expressed by some superficial measure, the average wage, the average income of persons counted for the income tax, the average account of a holder of a savings bank account, the average duration of disease, the average distance covered by a passenger or a ton of freight, the average tonnage and the mean cargo of a ship, the average amount of a postal money order, the average number of members of a club or of a cooperative society, the average business share of a member of a cooperative society, and so forth.

The progressive development of statistical science leads to the continuous opening of new fields to statistical observation; new quantities are investigated statistically and expressed by statistical series. Every such new series suggests the possibility of the computation of an average. At the same time the methods already in use are refined and new facts are registered. This new information enables us to dissect the series originating from the investigations from new points of view and to divide them into components which, in turn, may give rise to new averages.

In this connection we must also mention the frequent use of arithmetic means in graphic representation. Since the arithmetic mean is obtained from a series by adding the items and by dividing the sum by the number of the

items, the sum of all the items is found by multiplying the arithmetic mean of a series by the number of its items. This fact explains the availability of the arithmetic mean for graphic representation in the form of rectangles. If we draw a rectangle, its base corresponding to the number of items and its height to their average size, then the area corresponds to the sum of all the items. Often different quantities are represented graphically in this way and then their special features are easily compared.

B. THE WEIGHTED ARITHMETIC MEAN

The weighted arithmetic mean ("das gewogene arithmetische Mittel," "moyenne arithmétique pesée," "composée" or "graduée," "media arithmetica ponderata" or "composta") does not represent an independent kind of mean. On the contrary it agrees in its essential qualities with the "simple" arithmetic mean and differs from it merely in one point of secondary importance.[23]

This difference is that in the computation of a weighted arithmetic mean the items are not simply added and the sum divided by the number of items, but that the items before their addition are multiplied by coefficients (weights)

[23] The mean which we call weighted arithmetic mean here, was formerly called geometric mean in Germany, a term with which we nowadays denote a kind of mean totally different from the weighted arithmetic mean, and which will be discussed in a later chapter. Haushofer still used the term geometric mean for the mean which in modern times is called weighted arithmetic mean. (Lehr- und Handbuch der Statistik, 2nd ed., 1882, p. 53. Such averages as were found with reference to the relative weights of the items of a series were called geometric, in opposition to the arithmetic, which were ascertained without reference to these weights.) Also G. v. Mayr called the weighted arithmetic mean geometric mean in his book, Die Gesetzmässigkeit im Gesellschaftsleben, p. 53, but in his Theoretische Statistik of the year 1895 he dropped this term and, following the English and Italian terminology, proposed the term "weighted mean," which has since been introduced generally.

of different sizes and the sum of the products resulting is finally divided, not by the number of items, but by the sum of all the coefficients. The fundamental principle of the computation of the two means is the same. However, when computing a weighted arithmetic mean, the series is first subjected to a change of formation, the purpose of which is to give to the single members of the series an influence varying with their importance, their weights.

The example of a weighted arithmetic mean usually quoted is the average price of a commodity with reference to the quantities sold at different prices. If 20 units, yards, pounds, tons, etc., of a commodity are sold at the price of 10, and 10 units at the price of 16, then a weighted arithmetic mean is computed from these data by first multiplying the prices by the quantities sold, adding these products [$(20\times10)+(10\times16)=360$] and dividing this sum by the number of the units sold [30]. In this manner we find the "weighted" arithmetic mean, 12. Without reference to the quantities sold we would obtain the average 13 from the two prices 10 and 16.

As a matter of fact we proceed in exactly the same way if we have wages and numbers of workmen, instead of prices and quantities of merchandise. When computing the arithmetic mean of wages the procedure is self-evident and nobody thinks of speaking of a "weighted" arithmetic mean. The series consists of 30 independent units (workmen). In order to simplify the series the items of equal value (equal wages of a number of the workmen) are not given individually. However, the number of workmen, who receive equal wages is known, and must be taken into consideration.[23a]

[23a] Scott Nearing defines the "simple mathematical [arithmetic?] average" erroneously in his Wages in the United States (Macmillan, 1911). He says (p. 120) that "the simple average, by far the least satisfactory, is secured by adding the rates of wages and dividing by the number of different groups of wage earners." This

On the other hand the price series, mentioned above as an example, does not consist of independently ascertained units. The pounds or yards, tons, etc., which have been sold, are merely arithmetic units that do not exist independently. But different quantities of merchandise were sold at different prices, and therefore the quoted prices have different weights. Consequently, it would be insufficient to merely add the different prices and to divide their sum by their number. It is evident that in the computation of the average the different weights of the single prices must be expressed. This is done by using as "weights" those quantities that were sold at the different prices. We pretend, so to speak, that the series consists of as many members as units of quantity sold and take every quoted price into account as often as quantity units were sold at that price. Since this, however, is really a pretense, we feel that we deviate from the general rule for the computation of an arithmetic mean and we call the mean computed with reference to the quantities sold, the "weighted," in contrast to the simple, arithmetic mean.

When computing an average price from a time series of prices we ought to proceed in a similar way as if the several prices were given for the same time. If in 10 successive years the quantities 1, 2, 3, 4, . . . 10 are sold at prices 1, 2, 3, 4, . . . 10, then the weighted average-price for the decade would be 7; i. e., $(1\times1)+(2\times2)+(3\times3)+(4\times4)+ \ldots (10\times10)$ divided by $55=7$. The "simple" arithmetic mean, which, however, would be incorrect, is $5\frac{1}{2}$ (the average of 1, 2, 3, 4, . . . 10).[24]

definition is at variance with the definitions given by A. L. Bowley (Elements of Statistics, p. 109) and G. U. Yule (Theory of Statistics, p. 108) as well as with the usage described above by the author.—TRANSLATOR.

[24] The principle mentioned is taken account of in the rules which regulate the procedure of the Austrian permanent commission for commercial values in the ascertainment of the annual average prices.

In the examples mentioned the point in question was to compute averages from items whose difference of weight was given numerically. In these cases the method of computation of a "weighted" arithmetic mean is more or less self-evident. Frequently, however, it is necessary to compute averages from items which evidently have different weight, although we have no numerical data which we could use as "weights." If, in such cases, we want to express the different importance of the items, then we are obliged to use estimated "weights." These weights must be chosen so that their relation to each other is proportionate to the surmised relation between the items.

Cases of this kind occur frequently. Statistical records are rare which state prices as well as the quantities of merchandise sold. The price lists of stock-exchanges merely quote the prices at which sales have been made on the different days, but not the quantities of stocks or merchandise sold at the different rates on the different days. Consequently the computation of an exact "weighted" average for a long period of time is impossible. If quantities of great difference were sold at the different prices, we could express this fact when computing the average for a longer period by the use of estimated "weights" proportionate to the surmised quantities sold.[25]

There it says, the commission must also take into consideration during what part of the year the greatest fluctuations of price have taken place and how the imports and exports of the whole year are distributed over the single parts of the year, i. e., the quantities imported and exported at various times during the year at the different prices must be taken into consideration in connection with the different price levels which have existed at such times.

[25] Average rates are needed in order to compute the revenues from bonds. Here annual average rates are mostly used as bases. In Austria, if customs are paid in silver (instead of in gold), a premium must be paid, the size of which is determined monthly according to the relation between the monthly average rate of the gold 20-franc

The direct estimation of "weights" is a last resort to be avoided if possible. If data for the calculation of the weights are not at hand we may, in order to avoid direct estimation, substitute numerical quantities which appear to approximate the true weights. The following example may be given: In order to compute accurately the average rate of interest of a bank for one year, it would really be necessary to combine the discount rates during the year with the loans that have been effected at such rates. This kind of computation is not practicable. In order to avoid a direct estimation of the weights of the different discount rates, we take a known measure which we surmise is proportionate to the weights of the items, i. e., to the amounts of loans. In computing the average rate of interest for the year we usually combine, therefore, the single discount rates with the length of time they have ruled, i. e., we multiply every discount rate by the number of weeks it was in force, add these products, and then divide the sum by 52, the number of weeks in a year. This assumption is, however, not free from objection. For the volume of loans is not always the same and depends to a great extent upon the rate of interest itself.

From series of measurements (wages, prices) the weights belonging to the various members of the series can be found in the majority of cases. The cases where this is not possible are usually occasioned by deficient information (for instance, if only prices and not the quantities sold are given). From series, however, which consist of relative numbers or means that refer to constituents of a larger totality (series of the third group) the weights belonging to the items can never be found directly. But we know that as a rule the items of such series have different weights. If these items are other than arithmetic means (medians or modes), then it is not appropriate to compute an arithmetic

pieces at the Vienna exchange and the monthly average rate of the coined silver.

mean of the items. But if the series consists of relative numbers or of arithmetic means, then the computation of an arithmetic mean from them is undoubtedly allowable. However, a weighted, rather than a simple arithmetic, mean must be found. Fortunately the difficult computation of such a mean from the items is usually not necessary. On the contrary, we are frequently able to find directly on the basis of the totality (to the constituents of which the items refer) that superior relative number or that superior arithmetic mean which represents the weighted arithmetic mean of the items.[26] If a series of death rates is given which refers to different parts of the country or to different groups of population, then the death rate for the whole country or for the entire population may be contrasted with these items as their weighted arithmetic mean. In a similar way the general average wage of all workmen represents the weighted arithmetic mean for the average wages of certain categories of workmen. Therefore, in series of relative numbers and arithmetic means of the third group the computation of a weighted arithmetic mean from the items is not necessary, if the general relative number or the general arithmetic mean is known as ascertainable. The average should not be computed from the items but rather from the fundamental data, on which the items themselves are based. The average should be computed from the items themselves only when the data necessary for its independent computation are lacking. The numerical size of the " weights " to be used is to be determined in every case with reference to all circumstances. The use of weights may be dispensed with only in those rare cases where the items have practically equal weights.[27]

The necessity for computing the average directly from items of different weights occurs if estimated averages are given as items—for instance, estimated average wages of agricultural laborers for the different sized sections of a

[26] P. 140. [27] P. 142.

country. Since individual measurements are not given, an average for a wider geographical territory, the whole country, cannot be computed on the basis of the totality of all individual wages. The average for the wider geographical territory can only be found on the basis of the given averages for the single sections. While doing this we must consider that the number of agricultural laborers varies according to the section and that therefore the average wages for the single sections are not of equal value. The number of laborers may, perhaps, be obtained from the data of a census of occupations or a census of agricultural establishments. If this is the case, correct weights are found. Otherwise appropriate weights must be estimated on the basis of other data.

The use of weights in the computation of mean index numbers to represent changes in the price level has caused much controversy. As is well known, the object of this computation is to obtain an average from the single index numbers which denote the prices of merchandise of a certain year as a percent of the prices of a standard year or period. This average enables us to compare all the prices of any year with the prices of the standard year or period, and thus the prices of different years with each other.

The ordinary arithmetic mean of the single index numbers seems to be insufficient, because in its computation the same importance, the same weight, is attributed to the price fluctuations of all commodities under consideration. Thus, a fluctuation in the price of a rather unimportant commodity has the same influence upon the numerical value of the mean index number as a fluctuation in the price of the most important commodity. Therefore, numerous modern authors have found it to be necessary to combine the indices for the single commodities with weights, in order to accentuate the different importance in commerce or in consumption. Some authors use coefficients chosen

at will which, according to their subjective judgment, seem to be proportionate to the importance of the various commodities. It is equivalent to the use of such coefficients if, in the computation of the average, several price quotations are taken for one commodity or if single commodities are quoted in different stages of production, as, for instance, Sauerbeck does. Instead of coefficients chosen at will we may also use coefficients that are computed on the basis of quantities numerically known or, at least, capable of estimation. Thus the Economic Section of the British Association has used the " estimated expenditure per annum on each article " as the weights of the single indices. In a similar manner Professor Conrad, in his works on price statistics, assigns to the various commodities weights proportionate to their consumption. The British Board of Trade obtains its mean index number by computing the value of the foreign trade of a certain year, first, on the basis of the prices of this year and then on the basis of the prices of the standard year by stating the former value as a percent of the latter. The quantity of the single commodity which has been sold in the foreign trade of that year is used as its weight. *Vice versa,* the value of the trade of the standard year can be computed at the prices of various other years and the results can then be compared with the value which the trade of the standard year shows at the prices of such standard year. In this computation the quantities of the single commodities handled in the foreign trade of the standard year are used as weights.

If mean index numbers are computed for a series of years, usually the same weights are used in the computation of all the total index numbers. Thus Professor Conrad, who considers the consumption of the various commodities to be a measure of their importance, uses the quantities consumed in the year 1880 as fixed weights for all former and later years. But we may also use weights which change in the course of years in the same

manner as the relative importance of the various commodities.[28]

Weights can also be employed in a similar manner if total index numbers are used in representing the change of other complex statistical data. Thus, Bowley has used weights in the computation of his mean index numbers for the changes of the wage level; Wood used them in computing indices to represent the changes of consumption in England.

Bowley combined the indices which represent the fluctuation of wages in single occupations into mean index numbers for more extensive groups of occupations and indices for different localities into mean index numbers for wider geographical territories and, in doing this, he used weights in order to allow for the varying importance of the occupations and localities.[29] He also tried, by the use of changing weights, to allow for the changes which have taken place in the course of time in the different occupations and localities.[30] Wood represented the quantities of various articles of food (flour, cocoa, coffee, meat, rice, sugar, tea, tobacco, etc.), consumed per capita of the population during the years 1860-1896, by means of index numbers as a per cent of the average consumption of these articles during the standard period 1870-1879, and computed from these single indices the simple arithmetic mean, and five kinds of weighted arithmetic means in order to allow for the impor-

[28] Although Bowley (Elements of Statistics, Chap. IX, "Index Numbers," p. 220) does call the prices of the standard year weights, it is an incorrect expression. The deviation of the prices of a certain year from the prices of the standard year naturally depends on the level of the latter, therefore it is important to choose a standard as normal as possible. But the height of the prices of the standard year has no influence upon the *manner of the computation* of the mean, but only upon the numerical size of the items from which the average is computed and, therefore, on the magnitude of the average.

[29] See, for instance, Journ. of the Roy. Stat. Soc., Vol. LXII (1899), especially p. 712, and Vol. LXIX (1906), p. 164 ff.

[30] See Journ. of the Roy Stat. Soc., Vol. LXIX (1906), p. 167 f.

tance of the different articles of food. He computed the weights used principally on the basis of a typical workman's budget given by Booth in *Life and Labor of the People of London.*[31]

Modern statistics offers several interesting illustrations in which a mean is computed from a series of items by combining them into a weighted mean index number, the weights of the items being determined by the peculiar purpose in view. The United States Weather Bureau has computed the average rainfall, weighted according to population. The average rainfall of a country, geographically considered, should be computed by attributing different importance to the different rainfalls on record according to the area covered by them. However, the purpose of the American statistics is to represent the average rainfall according to its importance to the population. For this purpose the various measurements of the rainfall are not weighted according to the areas covered by the different rains, but according to the population of the areas. If the importance of the rainfall to the population is to be represented, then rainfalls in uninhabited districts evidently need not be taken into account. The importance of rainfall varies with the density of population. In this sense the average rainfall in the United States has decreased from 42.5 in., in 1870, to 41.4 in., in 1890. But this does not prove that a meteorological or climatic change has occurred but is caused principally by the fact that the drier Western states have been settled.

We proceed in a similar way when computing the mortality index which is found for a certain population on the basis of the age and sex classification of a standard population. A weighted arithmetic mean is computed from the special death rates for the different classes of age or

[31] "Some Statistics Relating to Working Class Progress since 1860," Journ. of the Roy. Stat. Soc., Vol. LXII (1899), especially p. 655 ff.

sex of the population by combining these special death rates with weights which belong to them according to the age and sex classification of the standard population. The purpose of the computation of mortality indices for different countries on the basis of the same standard population is, of course, to find indices which are comparable for all countries independent of the different age or sex constitution of the respective populations.[32]

A counterpart to the method of the standard population is the method of the standard mortality.[33] According to Professor von Bortkiewicz it is "the comparison between the number of deaths actually occurring and the number of deaths expected to occur according to a standard mortality." To the general death rate computed in the normal way which a certain population shows on the basis of its age constitution and the mortality conditions in the various age classes, is contrasted the mortality rate which the same population would show if in its various age classes those mortality conditions were prevalent that are found in the population chosen as a standard. This standard mortality rate with which the actual death rate of a country is to be compared, is found by computing a weighted arithmetic mean from the special death rates for the single age classes of the standard population. In this operation every one of these death rates is given that weight which belongs to corresponding age class according to the age constitution of the concrete population in question. This method plays an important rôle in the practice of life insurance com-

[32] Compare especially "Mortalitäts-Koeffizient und Mortalitäts-Index" in the Bull de l'Inst. intern. de Stat., Vol. VI, No. 2, and "Über die Berechnung eines internationalen Sterblichkeitsmasses (Mortality-Index)" in Conrad's Jahrbücher, 3rd series, Vol. VI (1893), by Josef Körösi, as well as the works of Ogle, Rubin, Sundbaerg, and v. Bortkiewicz.

[33] Compare Über die Methode der "Standard Population" by Dr. L. v. Bortkiewicz, Berlin, 1903 (Reports of the 9th Session of the International Statistical Institute).

panies, but so far has found little use in general population statistics.[34]

Certain statisticians have noted that, in certain cases, the use of weights has only very insignificant influence upon the numerical value of the arithmetic mean, so that, with or without the use of weights, and also with the use of different weights, we often obtain almost the same mean. This observation was made especially in the computation of weighted arithmetic means from series of price indices and it has been discussed by Giffen, Sauerbeck, Taussig, and others. The use of weights in the computation of mean index numbers, which is a laborious process and continually leads to controversies, appeared to be superfluous, and the computation of simple arithmetic means from indices of prices seemed to be justified.[35]

However, experience gathered from isolated cases and under certain conditions must not be immediately generalized. If only a few items of very different weights are given, then, of course, we obtain decidedly different values for simple and weighted arithmetic means. In such a case we certainly cannot do without the use of weights. But if a large number of items is given, then it is quite possible

[34] The method of the expected deaths—as the most important application of the more general "Method of expected events"—was advocated especially by Westergaard (cf. his "Alte und neue Messungsvorschläge in der Statistik"), Conrad's Jahrbücher, 3rd series, Vol. VI (1893), p. 330 ff. Bleicher has also concerned himself with this method (cf. v. Mayr, Bevölkerungsstatistik, p. 220).

[35] In his work in the field of historical wage statistics Bowley has found that weights eventually have only small influence on the size of the arithmetic mean (cf. Economic Journal, Vol. V, p. 373, and Journ. of the Roy. Stat. Soc., Vol. LXII (1899), p. 712, and Vol. LXIX (1906), p. 164 ff.). Wood obtained a similar result in his investigations on the development of English consumption, where he computed and compared simple arithmetic means and arithmetic means weighted according to 5 different systems (cf. "Some Statistics Relating to Working Class Progress since 1860," Journ. of the Roy. Stat. Soc., Vol. LXII (1899), particularly p. 655 ff.).

that the weights of the measurements above the average approximately balance the ones below the average.[35a] The use of weights in such a case is without effect, since the weights neutralize each other in the computation of the mean. If there is no relationship between the numerical size of the items and the weights which belong to them, then it is probable that the weights of items above and below the average are approximately equal and thus have no noticeable effect. In this way the fact that weights frequently have hardly any influence upon the arithmetic mean may be explained.[36]

However, the weights of the items must not be neglected if there exists a close connection between the numerical value and the weight of the items. Bowley gives a striking example of this. Suppose we desire to compute the average wage for a whole country from the rates paid to workmen of a certain occupation in the different cities of the country. We may take the rates found in the different cities to be of equal weight and compute a simple arithmetic mean. But we may also take into account that the number of workmen belonging to the occupation is different in the different cities, and use the corresponding numbers as weights in the computation of the mean. The latter method of computation will result in a considerably higher average number,

[35a] Thus, the United States Labor Bureau used the simple average in computing the general index number of wholesale prices of some 250 commodities and the Canadian Department of Labor likewise used the simple average of the indices of wholesale prices of 230 articles. In computing an index of retail prices the United States Bureau of Labor uses both the simple average of thirty articles of food and the weighed average in which the weights are chosen according to the average family consumption as shown in 2,567 budgets. The average absolute difference between the simple and weighted averages for the years 1890-1906 is 0.33; the difference exceeds 0.6 in but one year, 1900, when it is 1.4. (See Bulletin of the Bureau of Labor No. 77 for retail prices, 1890 to 1907, and Bulletin No. 87 for wholesale prices, 1890-1910.)—TRANSLATOR.

[36] Bowley, Elements of Statistics, 2nd ed., p. 117.

as there is a close connection between the amount of wages and the number of workmen. In the larger cities, with a greater number of workmen belonging to an occupation, the wages are generally higher and hence the weighted and simple arithmetic means must differ considerably.

The question, whether weights are to be used or not, cannot, therefore, be decided in general. The answer depends on the question of the relationship between the items and their weights. This latter question, however, cannot always be decided *a priori*. Therefore it is sometimes necessary to experiment with weights. If by using them we find a value which does not differ essentially from the simple arithmetic mean, then the use of weights evidently has no practical importance. But if the weighted arithmetic mean differs considerably from the simple arithmetic mean, then as a rule the consideration of the weight of the items cannot be neglected.[36a]

C. THE ARITHMETIC MEAN AND MATHEMATICAL STATISTICS

The arithmetic mean is widely used in the theory of error and the theory of probability, the principles of which

[36a] A committee of the British Association was appointed to investigate the question of price index numbers. Sir Robert Giffen states the conclusion of the committee as follows (Report of the Brit. Assoc., 1888, p. 184; also quoted in article on "Index Numbers" in Palgrave's Dict. Pol. Econ.): "The articles as to which records of prices are obtainable being themselves only a portion of the whole, nearly as good a final result may apparently be arrived at by a selection without bias, according to no better principle than accessibility of record, as by a careful attention to weighting. . . . Practically the committee would recommend the use of a weighted index number of some kind, as, on the whole, commanding more confidence. . . . A weighted index number, in one respect, is almost an unnecessary precaution to secure accuracy, though, on the whole, the Committee recommended it."—TRANSLATOR.

have been applied to statistical series and means.[37] The most important of these principles will be given as briefly as possible, with reference to the literature, and the consequences arising from the application of these principles to statistical data will be pointed out. While doing this we must distinguish between series of quantitative individual observations and series of statistical items, each of which expresses some numerical probability.

1. THE ARITHMETIC MEAN OF SERIES OF QUANTITATIVE INDIVIDUAL OBSERVATIONS AND THE THEORY OF ERRORS OF OBSERVATION

In order to judge the reliability of the arithmetic mean of a series of quantitative individual observations from the standpoint of mathematical statistics, we must go back to the theory of errors of observation. It is a fact, based upon experience, that repeated measurements of an object, the size of which is to be found, do not completely coincide. The reason for this lies in the fact that neither the human senses nor the measuring instruments are absolutely accurate. Therefore, the individual measurements are affected with accidental errors of observation. These accidental

[37] As the principal representatives of mathematical statistics may be mentioned: Lexis, Edgeworth, v. Bortkiewicz, Westergaard, Galton, Pearson, Yule, Bowley, Fechner, Czuber, and Blaschke. J. v. Kries has extended the application of the calculus of probability to statistics, especially on the logical and the perceptive theoretical sides. G. F. Knapp is the most prominent opponent of the application of the calculus of probability to statistical investigation (cf. his articles, "Die neueren Ansichten über Moralstatistik" and "Quetelet als Theoretiker," in Conrad's Jahrbücher, Vols. XVI-XVIII, 1871-1872). A. M. Guerry also expressed himself against the use of the calculus of probability in statistics (Statistique morale de l'Angleterre comparée avec la statistique morale de la France, Paris, 1864, p. xxxiii ff.).

errors are partly positive, partly negative, and we know from experience that smaller accidental errors occur more frequently than larger ones. Moreover, according to the theory of errors of observation, numerous measurements of the same object group themselves around their arithmetic mean according to the Gaussian law of accidental errors, and this mean may be considered to be the most probable value of the quantity measured. The true value of the quantity under observation cannot be determined. But superior and inferior limits can be found, between which the true value lies with a given numerical probability. These limits may be made to approach each other in two ways, first by procuring more accurate instruments, since the accuracy of the arithmetic mean of a number of observations depends upon the accuracy of the single observations, second by increasing the number of observations, for the precision of the arithmetic mean of a number of measurements varies directly with the square root of the number of measurements. Therefore, in order to obtain a result twice or three times as precise, we have to make four or nine times as many observations. The greater the number of observations, the greater will be the accuracy of their arithmetic mean and, therefore, the greater the probability that the true value is between given limits above and below this mean. Or, stating the same idea differently, the greater the number of observations the nearer together are the limits between which the true value is contained with *given* probability. With an infinite number of observations their arithmetic mean ought to represent the true value of the measured quantity. The accuracy of the determination of the arithmetic mean, called its "precision," may be expressed numerically according to a certain formula. The reciprocal of the precision, the "modulus" (the square of which is called the "fluctuation"), the mean error, and the probable error of the mean may all be used as measures of accuracy. They are con-

nected by mathematical formulæ so that one may be computed from the other.[37a]

[37a] Suppose that a very large number of measurements of a *single* physical quality are taken. Suppose further that our measuring instrument is so adjusted that there is no uniform tendency to give too large or too small values; in other words, to give *systematic* errors. However, there will be variations in the resulting measurements; the measurements will be affected by *accidental* errors. Suppose that we make the following assumptions:

(i) That the arithmetic mean of the observations is the most probable value of the quality that is being measured;

(ii) That positive and negative errors are equally probable;

(iii) That small errors are relatively more probable than large ones;

(iv) And that the contributory causes of error are independent.

If we let x stand for the accidental error (the difference between any observation and the arithmetic mean of all observations) then the law of distribution of errors or, in other words, the frequency-curve of error is:

$$f(x) = \frac{h}{\sqrt{\pi}} \cdot e^{-h^2x^2}$$

where

$e = 2.71828 \ldots$

$\pi = 3.14159 \ldots$

$h = \sqrt{\frac{n}{2\Sigma x^2}} =$ *the precision.*

$n =$ number of observations.

$\Sigma =$ "sum of such terms as"

$\frac{1}{h} =$ *modulus.*

$\frac{0.4769363}{h} =$ *probable error.*

$\frac{\Sigma|x|}{n} = \frac{1}{h\sqrt{\pi}} =$ *mean error.*

($|x|$ indicates that all errors are considered positive.)

$\frac{1}{h^2} =$ *fluctuation.*

—TRANSLATOR.

Therefore, the arithmetic mean of a series of repeated measurements of the same object has a special scientific importance in the theory of errors. It is only by computing the arithmetic mean from the measurements that the true value of the measured object which is of primary importance, can be closely approximated. Every single measurement may show an error of unknown quantity. The theory of errors is of especial value in astronomy and geodetic survey. In order to find the true size of an object, arithmetic means are computed from numerous measurements of it.

In statistics it is not a question of repeated measurements of the same object, but of measurements of different similar objects, each being measured but once. Now it has been noticed that statistical series sometimes show the same distribution around their arithmetic mean as repeated observations of the same object. This distribution, which corresponds to the Gaussian law, is especially characterized by the facts that the single observations are grouped symmetrically around their arithmetic mean, and that the items are densest around the arithmetic mean and become rarer the farther they deviate from it. On account of the concentration and symmetrical distribution of the items around their arithmetic mean, the mode and the median of the series theoretically coincide. Practically, however, they usually differ somewhat because of the comparatively small number of observations at hand.

A statistical series, the items of which are distributed around the arithmetic mean according to the Gaussian law, has the appearance of a series of repeated measurements of the same object. It is an obvious step, therefore, to apply the theorems resulting from the theory of errors of observation to such statistical series and their means. Mathematical statistics first investigates the distribution of the items of statistical series around the arithmetic mean. In case series follow the Gaussian law they are called

"typical" series, and the means from such series "typical" means. If a "typical" arithmetic mean is given, its accuracy is ascertained and the dispersion of the series around the mean is measured in the same way as though it were a question of mere errors of observation with repeated measurements of one and the same object.

Obviously, the ideas of the theory of errors cannot be applied to typical statistical means without certain changes. The arithmetic mean of a series of measurements of different objects (measurements of human height, length of life, wages) cannot be considered to be the most probable value of the "true" size of a certain object, since all the single units of observation are independent real phenomena and equally "true." However, we may take the arithmetic mean of the measurements to be a normal value (or the most probable empirical determination of a theoretical normal value) the size of which is determined by the general common causes influencing all the units of observation and from which individual cases differ merely on account of the disturbance due to individual accidental causes. Accordingly the arithmetic mean represents the "type" of the observed phenomenon which is expressed with merely accidental variations in the individual cases. From this argument the great scientific importance which mathematical statisticians attribute to "typical" averages can be recognized. The typical means are independent scientific perceptions. The series of items compared with the typical mean thus loses the greatest part of its importance. It is worthy of consideration only as a measurement of the variability of the type.

The application of the theory of errors of observation to statistical series of measurements undoubtedly has great theoretical value. Its practical importance, however, is small, since statistical series of measurements which conform to the Gaussian curve occur only very rarely. A number of such series were found in anthropometry, and

Quetelet thought that anthropometric measurements, especially height, were generally distributed symmetrically around their arithmetic mean according to the law of errors. More recent investigations, however, have established other forms of distribution in most cases. If statistical series show any regularity at all, it agrees only in very rare exceptions with the normal Gaussian curve. As a rule regular conformations of a different kind occur, such as the asymmetrical Gaussian curve emphasized by Fechner, or the skew curve of error. But in such cases the application of the principles of the theory of errors of observation to statistics loses a great deal of its importance. This theory cannot be applied to means of series that do not follow the Gaussian law. In such series arithmetic mean, mode, and median are separate values and often differ considerably, and there is no " normal value," in the strictly mathematical sense, from which the single items differ only by accidental deviations. There is no " typical " mean which could be used to stand for the series as a whole. But still the arithmetic means of series which do not coincide with the normal Gaussian curve are by no means inadmissible or unimportant. On the contrary the arithmetic mean offers valuable information about the series in question, even if it lacks the special justification found in the theory of errors. It characterizes the series, as has been shown above,[38] in many important points.

The mathematical-statistical argument is based on an analogy between statistical series of once-occurring measurements of separate but similar objects and repeated measurements of the same object, customary in geodetics and astronomy. In this the deviations from the average which the items of a statistical series (human heights or lengths of life) show, are placed on a par with " accidental " errors of observation and are treated according to

[38] See the section " Concept and Qualities of the Arithmetic Mean," pp. 138 ff.

the mathematical theorems developed for their treatment.

The principles of the theory of errors, however, may be applied to series of statistical measurements in still another way. In this application we do not place the deviations from the average on a par with errors of observation. On the contrary, we examine the actual errors of observation of each of the items and investigate the connection existing between these errors and the error of the average. For errors of observation are connected not only with astronomical measurements but, in most cases, also with statistical measurements.

The errors which occur in statistical measurements may be like the errors with repeated measurements of the same object, either systematic errors (biassed errors) or accidental errors (unbiassed errors). Systematic errors are all made in the same direction. If the construction of a physical instrument is incorrect, then all the single measurements show the same systematic error, and the error appears in the average. Systematic errors frequently occur in statistics. If many women, out of vanity, have stated their ages too low in the census, then this is a systematic error adhering to the age data which must be reflected in the average age of women. The error of the average age is equal to the average error of the items. A systematic error will also occur in an investigation of supposedly representative cases (not a complete census) if the choice of cases is ruled by criteria which tend to make them deviate in the same direction. For instance, wage data ascertained in a representative investigation often come mainly from the larger and more prominent factories where higher wages may be paid than in smaller factories. The average wage evidently must be higher than if all the factories had been taken into account uniformly.

Accidental errors, as distinguished from systematic, are, as a rule, partly positive, partly negative. They form the subject-matter of the theory of errors. The arithmetic

mean possesses the property that in its computation the accidental errors of observation adhering to the single observations neutralize each other, the more completely the greater the number of observations. This property originally proven for repeated observations of the same object also holds for errors connected with single observations of different but similar units. Also in this case the error of the average is considerably smaller than the probable error of a single observation, and the accuracy of the average varies directly with the square root of the number of observations.[39] The median and mode have no similar property. These means are obtained by selecting certain items as being characteristic of the whole series and, therefore, they share the errors of observation of concrete items.

2. THE ARITHMETIC MEAN OF SERIES OF STATISTICAL PROBABILITIES AND THE THEORY OF PROBABILITY

Just as the theory of errors of observation, under definite conditions, gives a reason for using the arithmetic mean of statistical measurements, so the theory of probability, likewise under definite conditions, provides a reason for adopting the arithmetic mean of series of statistical probabilities. The reasoning of mathematical statisticians, which is shortly explained in the following, is based upon the law of great numbers formulated by Bernoulli and Poisson, mainly with regard to experiences in games of chance, or rather upon the inversion of this law (theorem of Bayes).

If a certain theoretical probability exists that an event is going to happen, then an approximation to this probabil-

[39] Bowley, especially, has examined carefully the connection between the accuracy of a statistical average and the accuracy of the items on which the former is based. (Cf. "Relations between the Accuracy of an Average and That of Its Constituent Parts" in Journ. of the Roy. Stat. Soc., Vol. LX (1897), p. 855 ff., and Elements of Statistics, Chap. VIII, "Accuracy.")

ity may be found by experiment. That is, the empirical probability originating from experiment usually agrees approximately, but as a rule not completely, with its theoretical probability. If by continued experiment several empirical probabilities (i. e., empirical values affected with merely accidental errors) are obtained for the same theoretical probability, then these fluctuate symmetrically within certain limits around such theoretical probability. The arithmetic mean of the empirical probabilities is the nearest approximation to the theoretical probability and may be considered to be its most probable value.

The relation existing between the empirical frequency of an event and its theoretical probability obeys the law of great numbers. The greater the number of observations upon which the empirical frequency is based the more closely does this frequency follow the theoretical probability. The greater the number of experiments, the greater is the probability that the difference between the empirical frequency of the event in question and its theoretical probability is within assigned limits, or, stated in another way, the narrower are the limits within which the difference mentioned lies with assigned probability. The degree of accuracy with which an empirical value corresponds to the theoretical probability may, therefore, be expressed numerically in the individual case by the "precision" of the empirical value, or by the reciprocal of the precision, the "modulus," or by the mean, the average, or the probable error of the empirical value. If several values of an empirical probability are given for the same theoretical probability, then, according to the law of great numbers, their deviations from the theoretical probability vary inversely with the number of experiments on which they are based. In a given case, the probability of a given deviation from the theoretical probability can be computed.

Let us illustrate this by an example. If a drawing

is made from an urn containing red and white balls in the proportion 6:4, then there is a theoretical probability of 0.6 that a red ball will be drawn, a theoretical probability of 0.4 that a white ball will appear. Now let us make experiments by drawing a ball from the urn and putting it back 1,000 times in succession. It is very probable that the ball will not be drawn in the ratio of 600 to 400, but empirical probabilities for the red or white balls will be found which deviate slightly from the theoretical probabilities of both colors. If we repeat this process and ascertain the percentages of the red and the white balls for several successive drawings of 1,000 balls each, then we find for each color a series of empirical probabilities (i. e., empirical values, affected with merely accidental errors) which fluctuate symmetrically within certain limits around the theoretical probability. The theory of probability enables us to express numerically the degree of accuracy of the empirical values and to compute *a priori* the probability with which different deviations from the theoretical probability are to be expected. In two out of three drawings of 1,000 balls the number of the red would deviate at most 2.6% from the true number, i. e., the empirical value would be between 0.616 and 0.584, only very rarely there would occur a deviation in which the number of the red balls would be greater than 650 or smaller than 550.[40]

If we group, not 1,000, but 100,000 successive drawings from the urn, and if we find the percentage of the red and white balls drawn, then we obtain for each color an empirical probability which, according to the law of great numbers, probably more closely approaches the theoretical probability than an empirical probability based on only 1,000 drawings. Therefore, a series of values of empirical probability, each based on 100,000 drawings, will fluctuate around the theoretical probability within relatively nar-

[40] Cf. Harold Westergaard, Die Grundzüge der Theorie der Statistik, p. 57.

rower limits than a series of analogous values each based on only 1,000 drawings.

Thus, the law of great numbers explains what empirical values may be expected with assigned probabilities, if experiments are made which we base on a known theoretical probability. On the other hand, the inverse of this theorem permits us, under certain conditions, to draw conclusions from given observations to the unknown theoretical probability on which these observations are based. It is this inversion of Bernoulli's theorem which can be used in statistics. Statistical events show certain analogies to accidental events. The various statistical events (births, deaths, etc.) apparently occur with the same irregularity as the drawings of red and white balls from an urn. However, great regularity is found when many individual events are combined. Mathematical statisticians, therefore, frequently conceive relative numbers fulfilling certain conditions (statistical probabilities) to be empirical determinations of (unknown) theoretical probabilities or of functions of such.[41] From the empirical values conclusions are made as to the more important theoretical probability. It is possible to determine between what limits above and below the empirical value the theoretical probability is situated with any assigned probability.[42]

[41] Corresponding to the average character of most of the relative numbers (cf. above, p. 38 ff.) it can, as a rule, be assumed that they do not express a single uniform probability, but an "average probability," so that particular "special probabilities" exist for certain constituents. The consequences which result from the application of the theory of probability to statistical probabilities, have been treated thoroughly by L. v. Bortkiewicz in his paper, "Kritische Betrachtungen zur theoretischen Statistik, I. Artikel" (Conrad's Jahrbücher, 3rd series, Vol. VIII (1894), p. 641 f.). These consequences do not bear upon what is said in the following text.

[42] For this Czuber gives the following example (Wahrscheinlichkeitsrechnung, p. 304): of 54,391 males who completed the age of 50 years, according to the German mortality tables of 1883, 1,049

According to Bernoulli's theorem these limits are closer together the greater the number of observations. If we find, for instance, 2,600 male births in a total of 5,000, then the theoretical probability of the birth of a boy is determined only within rather widely separated limits. This relative number may indicate that the theoretical probability is located between $\frac{2700}{5000}$ and $\frac{2500}{5000}$. But if we have observed 500,000 births in which there are 260,000 males, then we may expect that the observed number differs only 1,000 at most from the theoretically most probable number; therefore this would be located between 259,000 and 261,000.[43]

The mathematical theorems mentioned evidently emphasize the importance of the arithmetic mean. If a series of relative numbers be given which can be considered to be empirical probabilities (for instance, a series of relative numbers which indicate the number of male births in proportion to the total number of births by years or districts), then the arithmetic mean of these numbers is based on a much larger number of observations than any individual relative number. Therefore, according to the theory of probability it approaches the true theoretical probability with which the mathematical statisticians are concerned, considerably nearer than do the individual relative members. It may be considered to be the most probable value of the theoretical probability. Thus the arithmetic mean possesses considerably greater scientific value than the items.

However, the application of the law of great numbers,

died before they reached the age of 51. The empirical value of $\frac{1049}{54391} = 0.01929$, with the probable error of 0.000397, is obtained from these data for the probability of death of the 50-year-old males, so that an even bet could be made that the probability mentioned falls between the limits 0.01889 and 0.01969.

[43] Westergaard, Die Grundzüge der Theorie der Statistik, p. 57 f.

or its inverse, to statistical relative numbers is feasible only under certain assumptions. The theorems of the theory of probability can only be applied to relative numbers which, in a formal way, can be considered to be probabilities or functions of such.[44] However, as is known, practical statistics has frequently to do with relative numbers that are not of such form. All the frequency numbers and coefficients which originate by correlating certain events (births, deaths, marriages, crimes, etc.) with the average population are neither probabilities nor functions of such. And yet these values (birth rates, death rates, marriage rates) form the chief material of practical statistics, while the corresponding probabilities—as is proven by the controversies concerning the determination of true probabilities of death—can only with difficulty be obtained in a mathematically correct way.

But even if the series in question consists of numerical probabilities the application of the theory of probability is, nevertheless, not without difficulty. Lexis and von Bortkiewicz assert that relative numbers can only be considered as empirical probabilities if they belong to a series of values which are grouped around their mean according to the theory of probability. Given such a series, which is called " typical " because the mean is typical, then it is natural to imagine that a theoretical probability exists which appears with merely accidental errors in the items of the series, which items usually refer to different geographical districts or periods of time. Now, we are entitled to compute the precision of the items as well as the precision of the mean. At all events we can assume that the latter is a closer approximation to the theoretical probability than any one of the items. However, series of relative numbers which not only obey certain formal conditions, but also show dispersion around a typical mean according

[44] What relative numbers these are has been explained above, pp. 19 ff.

to the theory of probability, are very rare in statistics.[45] Most statistical series do not show a sufficient coincidence with the theory of probability, so that that theory can only be applied in comparatively rare cases to the mathematical proof of the superiority of the arithmetic mean. The computation of the arithmetic mean from a series of relative numbers can only in very rare cases be justified by the assumption that the arithmetic mean is more valuable scientifically from the standpoint of the theory of probability than the individual relative numbers. Often the use of the mean arises because of practical reasons which call for a simplified result. Often the use of the mean is even undesirable, since characteristic differences which individual constituents exhibit are thus frequently obliterated.

The use of the calculus of probability for the determination of the degree of accuracy of a statistical relative number conceived as an empirical probability has been, therefore, to a great extent relegated to the background by the modern mathematical statisticians, especially by Lexis, while the older theoretical statisticians laid great stress on it. Lexis thinks that the chief advantages of the calculus of probability as applied to statistics are that it offers, first, a comprehensive scheme for frequency distributions and, second, a measure of the stability of statistical relative numbers.[46-47]

[45] This holds for series with "normal" as well as for series with "supra-normal" dispersion, of which the former—according to the explanation of Lexis—correspond to a constant probability, the latter to a probability subject to accidental fluctuations (cf. below, Part III, Chap. IV, C.).

[46] Cf. v. Bortkiewicz, "Die Theorie der Bevölkerungs- und Moralstatistik nach Lexis," Conrad's Jahrbücher, 3rd series, Vol. XXVII (1904), p. 247.

[47] The questions connected with those problems of the theory of probability are treated in the third part of the book, Chap. IV, C.

3. RELATION OF "MATHEMATICAL" TO "NON-MATHEMATICAL" STATISTICS; TYPICAL AND ATYPICAL ARITHMETIC MEANS

The theorems of the theories of errors and of probability constitute a mathematically precise statement of ideas which are familiar to non-mathematical statisticians and even to laymen lacking any knowledge of statistics. It is a matter of common knowledge that measurements of an object are usually affected with accidental errors; that an average from several measurements gives the true size of an object most correctly; and that the accuracy of the average increases with the number of measurements. It is obvious to every statistician that a single statistical value may differ greatly, on account of individual causes, from the "normal" quantity desired and that by computing an average from a greater number of homogeneous values disturbing causes are eliminated and a more reliable foundation is obtained.[48] It is likewise obvious that relative numbers

[48] This knowledge is the reason that, for various practical purposes, we are not satisfied with the number for a single year, as this number may be influenced by exceptional circumstances, but try to support our conclusions by the average results of several years. Thus, most of the mortality tables were not computed on the basis of the results of the census and the deaths of one year, but on the basis of the average of several censuses and of the average deaths of several years. In the computation of the German mortality table of the year 1887, the results of the census of each of the years 1871, 1875 and 1880 and the death rates of the period 1871-1881 were used. That the conditions of a single year cannot always be the standard is also taken into consideration in the different fields of legislation. Thus § 59 of the Austrian Public School law decrees that schools must be built wherever the number of school children that have to go to a school more than one mile distant averages above 40 for a period of 5 years. On the occasion of the Bosnian tax-census in the year 1905 it was decreed that the taxes of every community should be the average amount of the contributions of the community during the last 10 years. Many other examples could be given.

which are based on small numbers of observations are unreliable and that the accuracy of a ratio generally increases with the size of the quantities brought into relation.

However, this generally recognized connection between the number of observations and the precision of the average or the statistical probability cannot be stated exactly without using mathematical methods. The important question, When is a statistical quantity large enough to allow a conclusion? has been frequently raised by non-mathematical statisticians. No satisfactory general answer to this question has been given. As a method to enable us, in an individual case, to answer the question of the legitimacy of a conclusion with regard to the size of a quantity in question, the division of the observations into several sets has often been proposed. If these constituent sets result in values not greatly divergent, then the totality is great enough so that the law of great numbers operates, otherwise the totality is not great enough. According to the principles of the theory of errors or the theory of probability there is no certain limit when a totality may be considered to be great enough in general. The greater the totality, the greater the precision of the average or the probability computed for it. The constituents resulting from the division of a totality obviously exhibit divergent values, since they comprise fewer observations, and the averages and probabilities of the constituents fluctuate within definite limits around the average or the probability of the totality, these limits being rather far apart if the numbers of observations are small. Some non-mathematical statisticians think that the existence of sufficiently large quantities is proven if a series of values shows a regular formation—for instance, if a series of death probabilities increases or decreases in definite proportion with the age. But this regularity is never perfect and its degree also depends on the size of the quantities on which the items are based.

This discussion proves sufficiently the importance of the mathematical method for certain problems of statistics. On this question von Bortkiewicz expresses himself as follows: "It is a self-delusion to believe that we are able to work independently of the ideas of the theory of probability. In reality even the grimmest foe of the analogy with chance is ruled in his statistical work by conceptions that originate from that theory. Indeed, the scientific statistician daily asks himself the question, if in this or that case the amount of the material at hand is sufficient for accidents to neutralize or counterbalance each other. He applies, therefore, the theory of probability without his own will and knowledge and, consequently, in an unmethodological way, in the rough way of the pure empiricist."[49] Indeed, there is a danger in the fact that non-mathematical statisticians sometimes unconsciously apply certain theorems of the theories of errors and probability when their use is not warranted. In this class belongs principally the blind worship for great numbers—widely spread, especially in former times—and the desire originating from this feeling to combine great numbers of observations in order to compute an average or to obtain a relative number. If we consider the values combined in the light of the theories of errors and probability, then we find in many cases that with reference to the series of items we cannot assume at all that the value of the average or the relative number increases with the number of observations. The material with which practical statistics has to work really offers few opportunities for the direct application of the principles and methods of the theories of errors and probability. But the comparison of statistical material with these theories always offers an interesting standard for judging this material, and therefore is useful even if it occasionally results merely in the negative fact that certain

[49] "Die Theorie der Bevölkerungs- und Moralstatistik nach Lexis," Conrad's Jahrbücher, 3rd Series, Vol. XXVII (1904), p. 251 f.

theorems of the theories of errors and probability sometimes used intuitively by non-mathematical statisticians cannot be applied to the case at hand.

The differentiation between typical and non-typical series and means, made in mathematical as well as non-mathematical statistics, furnishes an especially interesting example of the common use of ideas drawn from the theories of error and probability. Mathematical statistics sees in a typical series one corresponding to the normal law of error or the theory of probability, a series " whose items are approximations affected by accidental errors, to a fixed base value." [50] A typical series naturally results in a typical mean, a non-typical series in a non-typical mean. In non-mathematical statistics, likewise, series and especially arithmetic means are divided into typical and non-typical groups (Bertillon, Block, Haushofer, etc.). Haushofer says: " The average may be a value which is approached by all the phenomena observed, with which they sometimes are almost identical. Then it is called a type of all the single phenomena." " But, on the other hand, the average may be merely an arithmetic abstraction, i. e., a value which, although having been computed from the members of a series, is not intimately connected with the single items. The average age of the population is an example. No definite line can be drawn between these two kinds of averages. The greater the differences of the single phenomena from which the averages have been computed, the closer the latter approach to mere arithmetic abstractions." [51] Series are classified in non-mathematical statistics as typical or non-typical, depending upon the arithmetic mean resulting from them, whether typical or non-typical.

[50] Lexis, Abhandlungen zur Theorie der Bevölkerungs- und Moralstatistik, VIII, "On the Theory of Stability of Statistical Series," p. 171.

[51] Lehr- und Handbuch der Statistik, 2nd ed., p. 53 f.

Non-mathematical statistics, as well as mathematical statistics, denotes, as typical averages, those which represent series in a peculiarly qualified way, and the series thus represented are called typical series. However, while mathematical statistics can use the theories of errors and probability as a measure to ascertain typical series and typical means with certainty, a similar precise and objective measure is lacking in non-mathematical statistics, and the decision whether a certain mean or a certain series may be called typical becomes more or less subjective. In general, non-mathematical statistics sees in a typical mean one which corresponds to a great number of items, and around which the whole series is distributed as symmetrically as possible and without very great deviations. Means not answering these conditions, i. e., means which lie outside of the main group of items or around which the items are not distributed symmetrically, are called non-typical. A definite line, however, cannot be drawn between typical and non-typical means in elementary statistics.

Since non-mathematical statistics, as has been mentioned, are lacking the objective criteria at the disposal of mathematical statisticians for the differentiation between typical and non-typical series and means, different conclusions may easily result in individual cases. A series of relative numbers and their mean may seem to be typical to the non-mathematical statistician, while the mathematician may not be able to find correspondence with the theory of probability; a series of measurements may appear to the former to be distributed with sufficient symmetry to denote a typical series and a typical mean, while the latter may find a contradiction to the Gaussian law. In general, the rules of the non-mathematical statisticians are less strict; this enables them to use the expression "typical" to a considerable extent, while the mathematical statisticians, as is known, have found but few series which might be called "typical" in their sense of the word.

But even if the idea of the " typical average " is defined in a manner not too rigorous, as is customary in non-mathematical statistics, even then decidedly non-typical averages occur only too frequently. These " mere arithmetic abstractions " can be used to represent a series only with great precaution. The best known example of such a non-typical average—besides the average age of the living obtained from the census figures—is the expectation of life, the arithmetic mean of the ages in the mortality table. In Prussia the mean expectation of life for the period 1881-1890 was 39 years, one month, in Austria 33 years and 8 months. In all countries it falls in the middle age classes which show relatively few deaths, while the majority of deaths occur in childhood and old age.[52]

[52] G. v. Mayr's use of the terms "typical" and "non-typical" means differs from the prevailing usage (Theoretische Statistik, p. 102). v. Mayr says: "A typical mean is given, if from the nature of the thing an ascertained average also represents a possible reality of the phenomena in question. This is the case with birth, death, and crime rates in time series, in absolute as well as in relative numbers. A merely arithmetic abstraction is given if an actual coincidence of all the objects with the average cannot reasonably be conceived. Such is the case with the ascertained average age of those living and those dying." A typical mean in v. Mayr's sense is the average of those marrying in contrast to the average age of those living. v. Mayr says in this connection (Bevölkerungsstatistik, p. 402): "The average age of the living population is merely an arithmetic abstraction; the existence of a population consisting only of people of average age is not conceivable. But it is not inconceivable that all those marrying do so always at the same age."

G. v. Mayr has also appropriated the term "typical series," contrary to the general statistical terminology, for a peculiar category of series which he has formed. He says (Theoretische Statistik, p. 90): "Typical series, in opposition to those series which offer only a concrete section of a continuous development, are those which from the nature of the observed material represent within themselves all the possibilities of a given phenomenon (for instance, distribution of a given population according to height, of a given number of births according to the number of children born in one confinement,

A remarkable change has occurred in the use of the word typical. By typical phenomena the earlier writers meant processes which are governed by certain laws of nature—for instance, most of the physical and chemical processes. It had been found that with such "typical" phenomena

and according to the sex of the children, distribution of births, deaths, crimes over the different seasons, etc.). With such typical series a natural arrangement of the members results from the quantitative or time graduation of the possibilities of the phenomenon." In the course of his discussion, however, v. Mayr approaches the conception defined by the mathematical statisticians and adopted also by the elementary statisticians in general. For v. Mayr continues: "The most pronounced form of such typical series seem to be those in which the items of a concrete series of observations must be considered to be quasi-inaccurate representations of an unchanging base-value, which is expressed in the phenomena actually observed only with purely accidental errors. The symmetrical arrangement of the items located below and above the mean characterizes this most pronounced form of typical series (frequency curves); the more asymmetrical this arrangement and the less decided a central elevation of the curve, the more does the series lose its typical character."

Adolphe and Jacques Bertillon, as well as Block, call non-typical means "moyennes indices" in opposition to "moyennes typiques," and the former expression also denotes certain isolated averages for potential measurements. J. Bertillon's principal examples for "moyennes indices" are the mean expectation of life and the average consumption of alcohol per capita of the population (cf. A. Bertillon, "La théorie des moyennes en Statistique," Journal de la Société de Statistique de Paris, 1876, p. 268, and J. Bertillon, Cours élémentaire de Statistique, p. 118 f., and Block, Traité théorique et pratique de Statistique, 2nd ed., p. 124). The astronomer Herschel, who wrote the preface to Quetelet's Physique sociale, proposed to denote non-typical means also in France by the English word average; but this word has not, in English, the meaning which Herschel attributes to it and it has found no place in the French language. Quetelet himself proposed to call typical means simply "moyennes," and to use the name "moyenne arithmétique" for non-typical means. This terminology was not accepted. A. Bertillon objected for the reason that it had not been chosen properly, since typical means are also computed in an arithmetical way.

the law established by the observation of a single case ought to hold without exception for all analogous cases. If a physicist has observed that a drop of mercury freezes at a certain temperature, then he can assume that this fact holds for all the drops of mercury in the world.[53] The older statisticians used to emphasize that such " typical " phenomena are not suitable objects for statistical research and that it is the province of statistics to investigate " individual " phenomena (i. e., those that vary from individual to individual), or as Meitzen [54] says: " To strive for the perception of the ' non-typical,' attainable only through the statistical method."

Furthermore, the older statisticians often made the mistake of confusing the antithesis between " typical " and " non-typical " phenomena with that between nature and society, because generally they considered the phenomena of nature to be " typical " and the phenomena of society to be " non-typical." [55] As a matter of fact, however, these distinctions do not coincide. It is true that social phenomena, as a rule, are non-typical.[56] They are not governed by one but by various causes of different importance, so that every phenomenon shows a more or less individual aspect. We cannot draw any conclusions as to the duration of life, the age of marriage, or income of specified in-

[53] Cf. Haushofer, Lehr- u. Handbuch der Statistik, 2nd ed., p. 38.

[54] Geschichte, Theorie, und Technik der Statistik, 2nd ed., p. 81.

[55] Cf., for instance, Rümelin, "Zur Theorie der Statistik," I (Reden und Aufsätze, 1875, p. 213 ff.).

[56] However, this rule is not without exceptions. In social life, and especially in economic life, there exist typical processes, which originate from a definite motive and, given the same conditions, always repeat themselves. Lexis (Theorie der Massenerscheinungen, pp. 2-4), in such cases, speaks of generic quantitative phenomena and gives the following example: If on the Berlin exchange the exchange rate on Paris goes above 81.40, then it may be asserted that all German bankers, who are at all prepared for operations of arbitrage, will send gold to Paris.

dividuals from the knowledge of those facts for other individuals. However, besides the " typical " phenomena, i. e., processes which according to the older terminology are governed by laws of nature with which statistics has nothing to do, nature also exhibits " non-typical," " individual " phenomena in great abundance, a fact which older statisticians do not seem to have observed sufficiently. Thus, dimensions of animals and plants are extremely variable, and therefore quantitative single observations are indispensable in biological research. Furthermore, meteorological phenomena, especially, are very changeable and consequently require quasi-statistical quantitative observations.

In modern statistics, the term " typical " does not serve to distinguish different categories of phenomena but principally to indicate definite series and means. The observation of those phenomena which the older statisticians called " non-typical " on account of their individual differences, results in statistical series and means which the modern statisticians divide into " typical " and " non-typical " according to the distribution of the items around the mean (thus following a criterion which is completely different from the criterion on which the older distinction between " typical " and " non-typical " was based).

4. MATHEMATICAL METHODS OF JUDGING THE SIGNIFICANCE OF THE DIFFERENCE BETWEEN TWO ARITHMETIC MEANS OR STATISTICAL PROBABILITIES

The principles of the theories of error and probability are frequently used by mathematical statisticians when comparing statistical averages or relative numbers which have the form of numerical probabilities. From the difference of two means or relative numbers, as shown in a previous chapter, we may, under certain circumstances, draw

inferences concerning the general causes acting on the phenomena compared.[57] While doing this we must distinguish between the comparison of quantities differing in geographical location and time, and those differing quantitatively or qualitatively. By comparing values which refer to different geographical or time divisions, we can ascertain the existence of differences, but we do not thereby determine the causes acting on the quantities in question. If we find, e. g., that the death rates of two countries differ, we can infer that the causal conditions influencing the mortality differ in the two countries, but for the time being we do not know wherein this difference lies. But if we compare masses differing quantitatively or qualitatively in a definite way and find a difference between the means or relative numbers ascertained for the masses in question, then we can, under defined conditions, ascertain immediately the cause of this difference. For if certain postulates are fulfilled we can then trace the difference of the means or relative numbers at hand back to the quantitative or qualitative difference between the masses on which these values are based. If we find that the death rate of males is higher than that of females, or that the death rates of those belonging to different occupations differ from each other, then we are justified, other things being equal, in attributing to the sex or the occupation a decisive influence on the mortality.

However, we can infer that different fundamental causes affect the items (having space, time, quantitative or qualitative differences) on which the compared means or relative numbers are based, only in case the difference between the compared values is significant. No inference can be drawn from very small differences. "One does not need a scientific training in order to be aware of this. It is evident that it is not a sign of improvement of sanitary conditions if in a certain locality 12,345 deaths occur in

[57] Cf. p. 110 ff.

one year and 12,344 deaths in the next year. The difference is so small that it may be accidental. But what does accidental mean in this connection? What are its limitations? What differences are great enough to pass from the realm of the accidental? If the answers to these questions are to be universal and not merely depend upon the idiosyncrasy of the statistician, which may lead the skeptical to an overestimation of the limits of the accidental and the sanguine to rash conclusions, then they must be based on the law of great numbers.''[58]

As a matter of fact, mathematical statisticians have developed special methods that enable them to ascertain the significance or lack of significance from the standpoint of the theories of error and probability of the numerical difference between two arithmetic means (means for an element of measurement), or between two relative numbers that may be assumed to represent empirical probabilities.

Mathematical statisticians consider means computed from elements of measurement to be mere approximations to the proper theoretical normal value of the quantity in question. This normal value is reflected in the concrete mean, but it is not expressed by it with perfect accuracy on account of the limited number of cases on which the mean is based. Therefore, different empirical means may correspond to the same theoretical normal value and, on the other hand, the same empirical mean may be based on different theoretical

[58] A. A. Tschuprow, Die Aufgaben der Theorie der Statistik, Jahrbücher für, etc., edited by Gustav Schmoller, 29th year (1905), 2nd number, p. 36. An example of judging a difference with and without the application of the theory of probability, is found in Edgeworth, "Methods of Statistics," Jubilee Volume of the Roy. Stat. Soc., p. 206. There Edgeworth asserts that in a certain case where Wappäus, in comparing data for two periods of time, had assumed a change in the mortality, a closer investigation by means of the theory of probability shows no sufficient reason for assuming a change other than accidental fluctuations.

normal values. Since, according to the opinion of mathematical statisticians, only the theoretical normal values (the accidental errors being eliminated) are determined by the general causes, therefore a difference in the fundamental causes operating upon two quantities can be inferred only in case a difference between the theoretical normal values of these quantities can be proven. Therefore the point in question is to ascertain if (and with what probability) the concrete empirical values, in spite of their numerical difference, are to be taken as approximations to the same theoretical normal value, or if, on account of this difference, they must be considered as approximations to different normal values. If it is very probable that the compared means are approximations to the same theoretical normal value, then the difference between the means is insignificant; if, however, the probability that the two means represent the same normal value, is small, then the difference is significant and an inference of different causation is allowable.

The practical application of the above theorems can be illustrated, without reproduction of mathematical details, by an example taken from an article by Professor Edgeworth. He compares the average height of 2,315 criminals with the average height of 8,585 persons of the normal population; the former average is 2 inches lower than the latter. Under the supposition that criminals have, in general, the same heights as the whole population, a modulus of 0.08 inch is found for the difference of two averages with the numbers of observation mentioned. If the actual difference between the two averages compared were not more than three times this modulus, then we could assume that this difference is merely accidental, i. e., caused merely by the small number of observations and by the accidental errors attached to these. But the actual difference (2 inches) is much larger than three times the modulus (0.24 inch); therefore, the difference is significant and

permits the inference that criminals are, on the average, of shorter stature than the normal population.[59]

This method of comparing averages may be used especially to ascertain periodical fluctuations. The differences existing between certain months of a single year are of course not conclusive; monthly averages must be computed and compared for a number of years. The method may also be applied to investigate whether or not a series shows a certain direction of development. For this purpose the averages resulting from successive periods are compared to ascertain if the differences between them must be considered significant or if they are merely accidental.[60]

The method of comparing statistical probabilities is based on considerations similar to those used above in the comparison of arithmetic means for an element of measurement. The probability is ascertained with which the two numbers compared may be considered, with regard to the numerical difference existing between them, to be empirical values of the same theoretical probability. If this probability is great, then the difference between the two numbers is insignificant and must be due to accidental causes. If, however, there is but slight probability that the numbers compared are empirical values of the same theoretical probability, then the difference between these numbers is significant, and it must be assumed that the phenomena compared reflect different theoretical probabilities and it fol-

[59] Cf. "Methods of Statistics," Jubilee Volume of the Royal Statistical Society (pp. 187 f. and 195 f.), where numerous other examples of the method under discussion are given. A mathematical method similar to Edgeworth's is used by Westergaard (Grundzüge der Theorie der Statistik, p. 187), who evaluates the actual difference by using the mean error of the difference of the two averages. See also the discussions by v. Bortkiewicz, "Kritische Betrachtungen zur Theoretischen Statistik," Conrad's Jahrb., 3rd series, Vol. X (1895), a, p. 334-341.

[60] See Bowley, Elements of Statistics, 2nd ed., p. 313 ff.

lows that the general conditions, which determine the theoretical probabilities, are different.

An example of the practical application of these ideas is found in the paper of A. A. Tschuprow, "*Die Aufgaben der Theorie der Statistik*" (p. 37 f.). Of 1,558,129 inhabitants 33,181 died in Vienna in 1897; therefore the empirical mortality amounts to 21°/₀₀; the modulus equals 0.17°/₀₀; therefore we may assume that the theoretical probability of dying is within the limits of 21.51°/₀₀ and 20.49°/₀₀. In Prague, in the same year, 6,392 persons died out of a population of 193,097; the mortality is 31°/₀₀, the modulus 0 56°/₀₀, the probability of dying, consequently, must lie within the limits of 29.32°/₀₀ and 32.68°/₀₀. Now since the inferior limit for Prague (29.32°/₀₀) is considerably higher than the superior limit for Vienna (21.51°/₀₀) it may be asserted that the two probabilities are different, and that the conditions of life with reference to the mortality are actually more unfavorable in Prague than in Vienna. The real cause of this, whether it is bad housing conditions, or impure drinking water, or unsanitary occupations of the population, or, perhaps, a difference of the sex and age constitution of the population, we do not learn for the time being it is true, but by ascertaining that the conditions of life in Vienna and Prague with reference to the mortality are not the same we gain firm ground for further research.[61]

[61] For the evaluation of the difference between two empirical probabilities various other methods may be used, for instance the method of comparing the actual difference with the modulus of the difference of frequencies (Tschuprow, loc. cit., p. 38), the method of comparing the actual difference with the mean error of the two frequencies or with the mean error of the difference of the frequencies (Westergaard, Die Grundzuge, p. 45 f. and p. 81 f., and Die Lehre von der Mortalitat und Morbilitit, p. 189 f.) and the method of comparing the actual difference with the probable deviation computed for the same (Lexis, Abhandlungen, "The Typical

Besides geographical comparisons we can, of course, also make comparisons of different time periods, classes of population, etc., in the same way. We may investigate whether a phenomenon shows annual fluctuations, or a certain direction of development, etc.

From a significant difference between two statistical probability figures we may thus infer the difference of the general causes acting on the phenomena compared. It is true the influence of the ascertained difference of causes cannot be determined with numerical accuracy, since the difference between the probability figures does not express it clearly, as this difference is also influenced simultaneously by accidental errors. However, it is—as Westergaard, particularly, emphasizes—of no essential importance to be able to express numerically the difference of the general causes; the main object is simply to ascertain that such a difference of the general causes exists.[62]

The method of the mathematical evaluation of the differ-

Values and the Law of Error," p. 128). See also Czuber (Wahrscheinlichkeitsrechnung, No. 165, "Probability That Two Empirical Determinations are Based on Unequal Statistical Probabilities," pp. 304-307). Czuber compares, for instance, the probability of a male birth among legitimate living births and legitimate still-births and finds that it can be asserted almost with absolute certainty, that with still-births there is a greater probability for a male birth than with living births.

[62] It is true that a difference whose effect is smaller than the limits of the accidental deviations cannot be ascertained. Westergaard (Die Grundzüge, p. 58) gives the following example: "If the mortality for one class of population is about 1% higher than for the population in general, then with a series of observations of 100 deaths, with a mean error of 10, we will never be able to assert the influence of a special cause, for the mean error is many times greater than the average effect of this cause. If 10,000 deaths were given, then we could assume such a cause with somewhat greater accuracy, but its effect would, on the average, not be greater than the mean error; but with 1,000,000 deaths a surplus of 10,000 deaths would betray a cause, the effect of which would surpass the mean error ten times and, therefore, could not pass as accidental. Here

ence between two relative numbers can be applied—as can the theory of probability in general—only to values which formally can be considered to be probabilities (or functions of probabilities). This means a considerable limitation of the applicability of this method since practical statistics much more frequently has to do with relative numbers which formally are not probabilities, than with relative numbers which fulfil the formal conditions in question. Furthermore, as has already been mentioned, it is not sufficient, according to the opinion of modern mathematical statisticians, that the two values to be compared are formal probability figures if taken by themselves. It must be an established fact that these values belong to groups of values which are distributed around their means according to the theory of probability. As is known, these conditions only very rarely hold true. The probabilities of death, for instance, compiled for a number of years, as a rule do not show a distribution corresponding to the theory of probability. Therefore, strictly speaking it is not legitimate to apply the theory of probability in comparing such items as probabilities of death of different classes of the population. Cases are, consequently, very rare in which the application of the theory of probability to the measurement of the significance of the difference between two relative numbers is entirely free from objection.

again the importance of a comprehensive series of observations is evident."

The above limitation also holds for the comparison of averages from data of measurement. Thus it cannot be ascertained by means of the theory of error, for instance, whether or not an additional duty, which is smaller than the mean error of the price fluctuations of the commodity in question, has had any effect on the domestic price.

CHAPTER III

THE GEOMETRIC MEAN

The geometric (or logarithmic) mean of n items is the nth root of their product. Where the items are represented by a_1, a_2 . . . a_n the formula for the geometric mean is $\sqrt[n]{a_1 a_2 --- a_n}$.[63] The great amount of arithmetic work involved in computing the geometric mean directly is lessened by the use of logarithms. The natural number corresponding to the arithmetic mean of the logarithms of a number of items is the geometric mean of those items.

The geometric mean has this property in common with the arithmetic mean, that in its computation the sizes of all the items are of decided influence on the size of the mean.[64] A change of a single item must affect the numerical size of the mean. This does not hold for the median or the mode. These means may remain unchanged even if considerable parts of the series are changed. The geometric mean has also this property in common with the arithmetic mean, that it may not coincide with any of the items used in computing it. As a rule the geometric mean is a value which does not occur in the series of items. If items of approximately the same size do not occur at all or only

[63] Therefore, by raising the geometric mean to the same power as there are items, the product of all these items is found, just as by multiplying the arithmetic mean with the number of items the sum of the latter is obtained.

[64] A series, in which an item equals zero, always gives zero for the geometric mean, without regard to the size of the other values of the series, since the multiplication of any number by zero gives the product zero.

rarely, we may call the geometric mean a mere arithmetic abstraction and we must consider it as an " atypical " mean.

The computation of the geometric mean like the computation of the simple arithmetic mean presupposes that the items have equal weights. If, in a concrete case, we think that this supposition does not correspond to fact, then we may treat this case by a method similar to that used in the computation of a weighted arithmetic mean from a series of items of unequal weights. Before computing the geometric mean we might modify the series by raising every item to the power which indicates its importance or weight and then find the nth root of the product of the rectified items, where n equals the sum total of the weights. In this way we might obtain, so to speak, a weighted geometric mean of the series.

The geometric mean plays a very subordinate rôle in practical statistics. It has been used by statisticians only sporadically, for instance, by Jevons in his monograph " A Serious Fall in the Value of Gold " (1863) for the computation of the mean index number from the single indices indicating the price fluctuation of various commodities.[65]

The geometric mean is never greater than the arithmetic mean of a series of items.[65a] The difference, however, is usu-

[65] See also the article "On the Variation of Prices," etc., by Jevons in the Journ. of the Roy. Stat. Soc. (1865), p. 294 ff. Jevons has not expressed the different importance of the different commodities, i.e., he has computed a simple, not a weighted, geometric mean.

[65a] To prove

$$\sqrt[n]{a_1 \cdot a_2 \cdot \cdots \cdot a_n} \leqq \frac{a_1 + a_2 + \cdots + a_n}{n}$$

The theorem follows by mathematical induction as follows (changing the notation):

(1) If a given quantity, a, be divided into three parts, x, y, z, the maximum value of the product xyz is attained when the

ally not very large, especially if the means are based on a great number of items. Thus, the geometric means of the 39 index numbers quoted by Jevons for the years 1851, 1853, 1855, 1857, and 1859 amount to 92.4, 111.3, 117.6, 128 8, and 116; the arithmetic means computed from the same index numbers for the same years are 94.6, 112.4, 119, 134, and 119.[66-67]

parts are equal, or when $x = y = z = \frac{a}{3}$. This theorem is easily proven by the method of the differential calculus.

(2) *Assume* that the maximum value of the continued product $x\ y\ z\ \ldots$ (to n—1 factors), where their sum is a constant, is attained when the factors are equal.

(3) Consider the product

$$w \cdot x \cdot y \cdot z \quad \ldots \ldots \ldots \quad (n \text{ factors}) = p$$
$$\text{where } w + x + y + z + \ldots \ldots = a$$

If any value, b, be assigned to w, then the maximum value of the product, p, is attained when the second composite factor is a maximum, or according to (2), when the n—1 factors are each equal to $\frac{a-b}{n-1}$. Consider, therefore, the function $w \cdot x^{n-1}$ and allow w to take values varying from o to a. For what value of w is this function a maximum?

$$p = w\,x^{n-1} = w \cdot \left(\frac{a-w}{n-1}\right)^{n-1}$$

for a maximum

$$\frac{dp}{dw} = \left(\frac{a-w}{n-1}\right)^{w-2} \cdot \left(\frac{a-nw}{n-1}\right) = o$$

$$w = \frac{a}{n}$$

or, the maximum value of the product $w \cdot x \cdot y \cdot z$ .... is attained when $w = x = y = \ldots\ldots\ldots\ldots = \frac{a}{n}$.

(4) But, since the maximum value of the product of three factors is attained when $x = y = z$, then, by (3), the maximum value of the product of four factors is attained when they are equal, and so on indefinitely to n factors.—TRANSLATOR.

[66] F. Y. Edgeworth, "A Defense of Index-Numbers," The Econ. Journ., Vol. VI (1896), p. 137.

[67] If we compute the arithmetic, the harmonic, and the geometric means of a set of items, then the last is at the same time the geometric mean between the first two means. The values 1 and 2,

Comparisons between the geometric and arithmetic means computed from the same series prove that the former is not influenced by extreme items to the same degree as the latter. The geometric mean of commodity index numbers is, therefore, less influenced by violent price fluctuations of single commodities than is the arithmetic mean. This property of the geometric mean is an advantage in such problems as in the representation of the movements of the price level where we do not think it justified for an exceptionally strong change in the price of a single commodity to influence the result to such a degree as is the case if the arithmetic mean is applied. Bowley recommends controlling the arithmetic mean by simultaneously computing the geometric mean of the items in question. If the arithmetic and geometric means of a series differ considerably from each other, the geometric mean must be considered to be more correct on account of the advantage mentioned above.[68]

The use of the geometric mean in computing mean index numbers has, as has been explained by Westergaard, the special advantage that the same result is obtained for a given period, no matter if this period is taken as a whole or divided into shorter epochs, which afterwards are combined. If the changes of the price level from 1860 to 1870 and from 1870 to 1880 have been computed by use of the geometric mean, then by combining these changes the same result is obtained for the price fluctuation from 1860 to 1880, as though the whole period had been treated

for instance, give the arithmetic mean 1.50, the harmonic mean 1.33, and the geometric mean 1.41. The last value is at the same time the geometric mean between 1.33 and 1.50. From this relation of the three means it follows that if two of them are given, the third may be computed directly from them (cf. Messedaglia, "Calcul des valeurs moyennes," Annales de démographie internationale (1880), p. 390).

[68] Elements of Statistics, 2nd ed., p. 128 f.

at once. This is not the case if the arithmetic mean is used.[69-69a]

[69] Cf. Westergaard, Die Grundzüge, p. 218 ff.; see also Bowley, Elements of Statistics, 2nd ed., p. 223.

[69a] Prof. A. W. Flux has tested the effect of a change of the base year with reference to which the commodity index numbers are calculated. The simple or weighted arithmetic mean of the commodity indices was found to vary by as much as 6% on account of the change. Of course a change of the base year does not affect the geometric mean ("Modes of Constructing Index-Numbers," Quar. Journ. Econs., Vol. XXI, p. 613.)—TRANSLATOR.

CHAPTER IV

THE MEDIAN

1. CONCEPT AND PROPERTIES OF THE MEDIAN

The median is that value which " has the central position in a series of items arranged according to size " (Czuber);[70] it is " the magnitude appertaining to the item halfway up the series " (Bowley).[71] Fechner defines the median (" Zentralwert ") as " that value of *a* "—*a* meaning the measurements of any collective object—" which has just as many values above it as there are below it and thus divides the series in the middle."[72-73] If an odd number of items arranged according to size is given, then the item in the middle of the series is the median; for instance, of 89 items arranged according to size the 45th is the median. With an even number of items the median lies between the two central items. It has the same size as they, if both are equally large; if they are not equal, the arithmetic average is usually taken as the median, or more accurate interpolation may be used.

The median differs essentially from the arithmetic and geometric means. In the computation of the last two the size of every item of the series is of influence since these

[70] Wahrscheinlichkeitsrechung, p. 334.

[71] Elements of Statistics, 2nd ed., p. 124.

[72] Kollektivmasslehre, p. 13.

[73] The values which, in a series arranged according to the sizes of items, form the line between the first and the second, and the third and the fourth quarters of the series are called quartiles; the values dividing the series into 100 parts are called percentiles. Quartiles, deciles, and percentiles may be used to supplement the median.

are added or multiplied in the process. The median is found in quite a different manner. The median is a definite item which, on account of its central position within the series, is considered to be characteristic of the series, or it is the average of the two items located in the middle of the series and therefore considered to be especially significant.[74] Changes of the items—except of the central member or the two central members—have no influence at all on the size of the median, as long as the *number* of items above and below remains unchanged. If a series consists of the numbers 1, 2, 3, 4, 5, 6, 7, 8, and 9, then 5 is their median. Any changes of the items above and below the median are entirely without effect, as long as the number 5 does not lose its central position in the series.

On account of its independence of single extreme cases the median may possess a more "typical" character than the arithmetic mean. If the arithmetic mean is used to represent the average income of a population, then the income of a millionaire will counterbalance the incomes of hundreds of workmen. The presence of a millionaire in a district otherwise poor will give an arithmetic average income which lies between the income of the mass of the population and that of the millionaire, a mere arithmetic abstraction, to which not a single real case corresponds. If, however, the median income is taken then the millionaire will have no more importance than any other individual and this average will undoubtedly reflect more accurately the income of the mass of the population.

However independent the median may be of extreme items, it depends entirely on the numerical size of the central item or items. Two series, in which only the central items differ considerably, while the other items coincide completely, result in very different medians.

[74] Liesse, therefore, correctly calls the median ("la médiane") "une moyenne de position" in opposition to the arithmetic mean (La Statistique, p. 82).

Changes of the central item or items of a series may cause considerable changes of the median, even if all the other items remain unchanged. The comparison of the medians alone produces, therefore, a one-sided picture in cases of the kind mentioned. These cases, however, rarely occur in practice.

The median is a typical value only if it appears at a point of concentration. In a series distributed symmetrically around a point of concentration it is identical with the arithmetic mean and the mode. If, however, the items are distributed in such a way that a concentration of the items takes place away from the center, then, of course, the median has no typical character. But the arithmetic mean of such a series would also be non-typical. In general, it may be said that most series can be characterized by the median just as well as by the arithmetic mean, while the former has the advantage that it is much easier to determine.[75]

It follows directly from the definition of the median that the numbers of items deviating in both directions are equal. Consequently, the probability that an item chosen at random lies below the median is just as great as the probability that it lies above. Therefore the median is sometimes called the " probable " value of the element of observation.[76] Thus the median age of the mortality table for all ages is called the probable length of life; Lexis calls the median of the age constitution of the living the probable age;[77] Boeckh has called the duration of marriage expressed by the median of his table of durations of marriage, the probable duration of marriage; in the theory of error the median of the series of errors is called the probable error, etc. When calling the median the

[75] See Lexis, Zur Theorie der Massenerscheinungen, p. 35.

[76] See Czuber, Wahrscheinlichkeitsrechnung, p. 334, and Fechner, Kollektivmasslehre, p. 166.

[77] Zur Theorie der Massenerscheinungen, p. 36.

probable value, of course, it must not be understood to mean that this is the *most* probable value. The most probable value is the one occurring most frequently, that is, the mode.

Furthermore, Fechner has proved that the sum of the deviations of the items from the median is a minimum, i. e., smaller than the sum of the deviations of the items from any other value.[78]

Accordingly the median differs mathematically from the arithmetic mean in two points. While the *sums* of the positive and the negative deviations from the arithmetic mean are equal, the *numbers* of the positive and the negative deviations from the median are equal; while the sum of the *squares* of the deviations from the arithmetic mean is a minimum, the sum of the *first powers* of the deviations from the median is a minimum.[79]

2. SERIES IN WHICH THE MEDIAN CAN BE DETERMINED

The determination of the median is customary only in series of quantitative individual observations (individual data—for instance, series of wages and incomes, ages, etc., series in the first of our three groups), and indeed, its use

[78] G. Th. Fechner, "Über den Ausgangswert," etc., Abhandlungen of the Saxon Society of Sciences, Vol. XVIII, Mathematical-physical group, Vol. XI (1874); see also Lexis, Zur Theorie, etc., p. 35, and Bowley, Elements of Statistics, p. 126.

[79] Fechner, Kollektivmasslehre, p. 13. In another place Fechner gives the following example: The series of the following 7 values chosen at random 0, 2, 4, 6, 7, 8, and 8 gives 5 as arithmetic mean and 6 as median. The sum of the deviation from the median is 17, but it is more from any other value no matter whether it is taken from the series itself or assumed between any of the items of the series—for instance, from the number 5 (the arithmetic mean), it equals 18; from the number 5.5, 17.5; from the number 2, 23. The sum of the squares of the deviations is smallest from the number 5 (the arithmetic mean); in this case it is 58; from the number 6 (the median) it is 65. (See "Über den Ausgangswert," p. 20.)

is free from objection only in such series, as will be proved in the following. The determination of the median can take place only in series the items of which are arrayed according to size. The members of the series of the second and third groups are usually arranged from other points of view than according to size. These two groups of series contain time, space, qualitative, and quantitative series. In time series the items are naturally given in chronological order, in space series they are arranged from some geographical point of view, i. e., according to the geographical location of the districts in question. In qualitative series the items follow each other corresponding to the logical sequence of the divisions; for instance, values for different occupations are arranged according to the relations existing between the groups of occupations distinguished. Finally, in quantitative series the arrangement of the items depends on the gradations of the quantitative criterion used in the formation of the series, consequently the values which refer to different age classes, for example, are given in the order of these age classes. In order to determine the median of a series of the second or third group we would have, first, to arrange the items according to size; in order to accomplish this the fundamental criterion that leads to the formation of the series would have to be disregarded and the series, therefore, would be destroyed.

There is another objection against the determination of the median for a series of the third group, i. e., for series the members of which characterize constituents of a larger totality in some definite manner by relative numbers or means. The members of such series (for instance, death rates for different geographical districts or occupations) usually refer to constituents of different sizes and, consequently, of different weights. The weights of single members, however, cannot be ascertained from the series itself. It is true, we can arrange the items of such a series according to magnitude, but we cannot ascertain the actual center.

The actual center of the series is located below or above the central item according as the inferior or superior items refer to greater constituents and, consequently, are of greater weight.[80]

From series of means we cannot ascertain the magnitudes of the individual observations on which are based the means forming the series. Consequently we can ascertain neither the order of magnitude of the actual items nor their central member. If the arithmetic average wages or normal wages for the different districts of a country are given, it is impossible to ascertain from them the " central " wage for the whole country, i. e., that wage which divides the individual wages of the workmen of the whole country in two equal parts so that one half of the workmen earn more, the other half less. Of course the average or normal wages may be arranged according to magnitude and the central member in this series of means be ascertained. But in this way a median is obtained which itself is again an average. This median will not be apt to coincide with that median which would result from the series of the individual wages of the whole country and which alone could be considered to be the actual " central " wage according to the definition.

[80] Colajanni is one of the few authors who determine the median for series of relative numbers of the kind mentioned above. (See Manuale di Statistica teorica, p. 182.) In the place cited Colajanni determines the median in the series of Italian marriage rates for the years 1872-1896 arranged according to size. We consider this procedure to be theoretically inadmissible, since the marriage rates mentioned are of different weights because the Italian population has increased considerably during the years in question and thus the data for the later years should have greater weight. However, the differences of weight in time series are usually much smaller than in geographical series and in series for different groups of population, and therefore the determination of the median for time series is less objectionable than for series of the last two kinds mentioned.

3. DETERMINATION OF THE MEDIAN

As a rule the determination of the median is very simple. The central item of a series of an odd number of items can be found by merely counting the items. If an even number of items is given, it is sufficient for most practical purposes to simply take the average of the two central items for the median. If, with an even number of items, we want to determine the median accurately, we must consider the formation of the whole series and try to interpolate that value which, according to the structure of the series, under the supposition that we are dealing with a continuous variable, would fall between the two central items. Usually, graphical interpolation is most expedient; for greater accuracy the series may be treated algebraically.

The median of a series consisting of an even number of items suffers, according to Fechner, because of the "inherent uncertainty of its determination."[81] For every value between the two central items—which Fechner calls the "limiting values" of the median—corresponds to the criteria established for the median. With every value between these limiting values the number of deviations on both sides is equal and for every such value the sum of the deviations is a minimum, because, when the median is moved between the two limiting values, an increase of the deviations on one side is compensated by a corresponding decrease on the other side.

Since the numerical value of the median does not depend on the sizes of all the items, but merely on the sizes of the items located in the center of the series, the median may under certain conditions be determined for series in which part of the items are unknown. The conditions which must be fulfilled are, first, that the number of the items must be known whose individual sizes are not known, and second, it must be established that the sizes of the items

[81] "Über den Ausgangswert," p. 20 f.

are such (even though unknown) as to place them unquestionably above or below the median. If these conditions are fulfilled, then the median of the series can be determined without difficulty. Cases are not rare where this property of the median can be used to advantage. In general investigations of income and wages certain groups of the population can often not be included. The excluded population are usually the classes with the smallest income and the lowest wages (for instance, persons having no trade and only occasional sources of income). If the number of persons, about whose income or wages no data can be obtained, is known approximately, and if it is certain that these persons belong to the lowest income—or wage—class, then the computation of the median for the total population is not impeded by the fact that the individual incomes or wages of these lowest classes is not definitely known. It is sufficient that the number of persons belonging to these classes can be taken into consideration when computing the median of the series.[82]

As has been noted, the practical statistician frequently has to work with series in which the items are not given according to their exact sizes but in classes of a frequency table. As has been explained,[83] the limits of these classes may be fixed either in order to give equal frequencies (such as percentiles), or regardless of the frequencies. Classes of the latter kind usually comprise varying frequencies no matter whether they have equal or unequal breadth.

If a series is given which consists of an even number of classes, formed according to the method of the percentiles, then the median is at the line of division between the two central classes or it lies in the middle between the superior limit of the lower class and the inferior limit of the adjoining higher class. If an odd number of such classes

[82] See, in this connection, Bowley, Elements of Statistics, 2nd ed., p. 125.

[83] See pp. 84-92.

is given, then the class can immediately be ascertained in which the median must be located, but its numerical size cannot be ascertained from the series. The median apparently lies between the two limiting values of the central class; its exact value, however, can only be ascertained by an exhaustive study of the series. The same thing holds if a series consisting of classes of varying frequencies is given. The class in which the median must be located is easily found. Within the limits of this class, however, the more accurate location of the median must be ascertained.[84]

The accurate determination of the median within the limits of a class is easily accomplished, if the original material of the investigation in question is at hand and if the grouping of the items within the class under consideration can be examined. This, however, is not usually the case. Therefore, the statistician will be obliged to form a hypothesis as to the distribution of the items within the class and to determine the median on the basis of this hypothesis. The simplest hypothesis is that of uniform distribution of the items within the limits of the class in question. If we use this hypothesis in a series consisting of an odd number of classes of equal frequencies, then, in order to obtain the median, we have merely to compute the arithmetic average of the two limiting values of the

[84] Since the exact sizes of the items belonging to the classes that evidently do not contain the median have no influence on its numerical size, the median may be computed directly from a series in which the two end classes have no superior or inferior limit, respectively, while the computation of the arithmetic mean of such a series involves considerable difficulty. As has been mentioned (on p. 145 f.), if a series of wage data indicates how many workmen belong to each of the following wage classes: less than $10.00, $10 to $10.99, $11 to $11.99 .... $29 to $29.99, $30.00 and more, the median could be computed from this without difficulty, while the accurate computation of the arithmetic mean is bound to fail because the sizes of the items less than $10 and more than $30 are not known.

central class. If we have a series consisting of classes of varying frequencies, then, using the hypothesis mentioned, the median in the class under consideration is usually determined by dividing the breadth of the class in question in the ratio which the median is supposed to divide the items of that class. This method, however, is not always clear or free from objection, as the following example is supposed to show.

Let a series of wage data (for instance, weekly wages) for 99 workmen be given. The wage data are given in classes of $2 breadth each. In the class that contains the wages from $24 to $25.99 there are 10 workmen, i. e., the 48th to the 57th inclusive. The wage of the 50th workman is to be determined since this wage represents the median. The 50th workman is the 3rd of the 10 workmen belonging to this class. *Prima facie* it is an obvious assumption that his wage is $\frac{3}{10}$ of the breadth of the class higher than the wage which forms the inferior limit of the class. Three-tenths of the breadth gives 60 cents, and the median would be $24.60. But the following argument is just as obvious: In the class $24 to $25.99 there are 9 intervals between the 10 items belonging to this class. According to the hypothesis of uniform distribution these intervals must be considered equal. Therefore the median is separated from the lowest item by 2 and from the highest by 7 intervals. In order to determine its size the breadth of the class must be divided in the proportion of 2 : 7. Accordingly the median would be $\frac{2}{9}$ of the breadth of the class above the inferior limit of the class and would be $24.44. The inconsistency arises because of the different positions which can be assigned to the first and last items in the class.

We can distribute 10 values in a class of 200 cents breadth so that the first and the last values coincide with the limiting values of the class; so that the first item coincides with the inferior limit while the last value is as far

distant from the superior limit as are the items from each other; or, so that the last item coincides with the superior limit while the first item is as far distant from the inferior limit as are the items from each other. None of these three distributions seems to be free from objection. The first kind of distribution, if carried out in the adjoining classes, would give two items at each class limit. The second and third kinds of distribution do not correspond at all to the postulate of a uniform distribution within the classes. The most correct way of distributing the items uniformly is to assume that they occur at equal intervals even when this distribution is extended to the adjoining classes. To fulfil this condition the first and the last of the items belonging to the class must be removed from the class limits to a distance which corresponds to half the magnitude of the interval existing between the items belonging to the class. Consequently in our example the wages of the 10 workmen belonging to the class \$24 to \$25.99 ought to be distributed in such a way that the wages increase 20 cents from workman to workman and that the first workman of the class receives \$24.10 and the tenth \$25.90. According to this computation the wage of the third workman, and thus the median, is \$24.50.

The above principles for the computation of the median from classes of a frequency table must also be used if the median is to be computed from a series of measurements for an element varying continuously. For series of this kind, as closer investigation will show, always consist of classes, even though they seem to give all the individual measurements. Continuous elements of measurement—for instance, distance, area, weight, duration of time, etc.—cannot be ascertained with complete accuracy, and in statistical measurement we are, as a rule, satisfied with a still smaller degree of accuracy. Individual items, which coincide in statistical observation, would be found to be of different values if greater accuracy were used. Therefore, the smallest units

distinguished really represent classes. How to proceed in the computation of the median from data for a continuous element of measurement, may be explained by means of a series of measurements of stature, which are taken from Francis Galton in the *Report of the Anthropometric Committee* of the British Association (1881).[85]

The series in question gives the heights of 76 boys, 13-15 years old. Since the series consists of 76 members, the median lies between the 38th and 39th member. The 38th as well as the 39th member are 59 inches. The four members preceding the 38th and the member following the 39th are also 59 inches. Nevertheless, the median is not exactly 59 inches. Since the heights have been recorded only to the nearest quarter of an inch, all the sizes between $58\frac{7}{8}$ and $59\frac{1}{8}$ inches in the series evidently are recorded as 59 inches. Now we can assume that the 7 heights called 59 inches are uniformly distributed between $58\frac{7}{8}$ and $59\frac{1}{8}$. Under this supposition the 38th and the 39th members lie between 59 and $59\frac{1}{8}$; their accurate position and the position of the median depend on the placing of the last member, i. e., the 40th. If this be placed at the superior limit of the class then the 38th and the 39th members are $\frac{1}{24}$ and $\frac{2}{24}$ inch above 59 inches. But if the items are distributed so that the last of them is still a certain distance away from the superior limit, a distance which corresponds to half the size of the intervals between the other items, then the 38th and 39th members are $\frac{2}{56}$ and $\frac{4}{56}$ inch above 59 inches. The median—the average of these two members —is $59\frac{3}{56}$ inches.[86]

[85] These data given in Bowley's Elements of Statistics were used by Galton to explain the method of graphic determination of the median, of which we shall speak later.

[86] Fechner, too, has examined the case, in which the median "*must be looked for within the uniform succession of the values of a measurement class.*" (See "Über den Ausgangswert," p. 18 f.) Fechner lays down the rule, that the median must be computed so that the same value is found no matter from which end we start to count

However, the hypothesis of the uniform distribution of the items between certain limits is not always admissible. On the contrary, sometimes the structure of the entire series makes it very probable that the items belonging to a certain class are not distributed uniformly within the class. If the class under consideration is preceded by a place of concentration, starting from which the items become less frequent in both directions, then it may be assumed that the items of the class under consideration are more densely crowded in the part towards the place of congestion than in the opposite part. In estimating the median from such series we must resort to hypotheses which are better suited to the formation of the series than the hypothesis of the uniform distribution of the items. Usually, results can be obtained most readily by graphic interpolation. The most precise determination is accomplished by means of algebraic interpolation. In this, just as in the graphic interpolation, we start from the hypothesis that the structure of the series intimated by the known items holds also for the items which are not accurately

and interpolate. This self-evident rule has also been observed above. On the other hand, there are differences between the above discussion and Fechner's statements. Fechner tries to determine the size of the 13th value in a class, the limits of which are 2½ and 3½ and to which 16 values belong. To this purpose he adds 13-16 of the breadth of the class to the lower limit (2½) and subtracts 3-16 of this breadth from the superior limit. In both ways he finds the same value, 3.3125. This value corresponds to the 13th item only under the supposition of a distribution of the items so that the last item coincides with the superior limit while the first item lies as much above the inferior limit as the interval between the other items amounts to. Such a distribution, however, does not correspond to the idea of a symmetric distribution of the items in the whole class. In spite of this, Fechner's result was correct; for although an even number of items was given, Fechner did not take the average of the two central items but the lower of these two items for the median. (See also Fechner, Kollektivmasslehre, p. 168 f.)

known, i. e., that the sizes of the items not known accurately are such that they do not change the structure of the series based upon the known values.[87]

Statistical series, both complete series of individual observations and series consisting of classes, are frequently adjusted or graduated in order to show clearly their characteristic formations and to remove the more or less accidental irregularities. In the adjustment of a series a number of its items are more or less modified. Consequently if the central items of the series are changed the adjustment may affect the numerical size of the median. The median of the adjusted series may be different from the median of the unadjusted series. But great differences are not to be expected.[88]

A graphic method of determining the median (as well as the quartiles and deciles) of a series has been explained by Galton in the *Report of the Anthropometric Committee* of the British Association of the year 1881 (p. 247). He uses the data above mentioned giving the stature of 76 boys, 13 to 15 years old. According to Bowley's presentation,[89] Galton's graphic method is essentially the following: The axis of abscissas of the diagram is divided into equal parts which correspond to the units of the element of measurements (in the present case inches), and serves, as usual,

[87] An example of the computation of the median of a series consisting of classes by means of algebraic interpolation is found in Bowley, Elements of Statistics, 2nd ed., p. 252 f.

[88] It is a kind of a limited adjustment (proposed by Fechner for a number of cases in "Über den Ausgangswert," p. 23) if, with an odd number of items distributed irregularly around the median, we take the arithmetic mean of the three central items for the median instead of the item which according to its position is the actual central item; "generally we may feel assured of coming closer to the true median, which would be obtained from an infinite number of items, by using this method, than by using the item which is influenced by accidental errors."

[89] Elements of Statistics, 2nd ed., p. 127 f.

to represent the different measurements occurring in the series. To express the frequency of these different magnitudes ordinates are used, but in a way differing from the usual. For it is characteristic of Galton's method that the ordinates belonging to the different magnitudes are not all placed, as usual, vertically on a common basis (the axis of abscissas), but, when placing successive ordinates, he measures from a new base at the height of the upper end of the preceding ordinate. Thus, it results that the upper end of the ordinate of the last (highest) measurement occurring is just as many units on the axis of ordinates above the axis of abscissas of the diagram, as the series has items. Each ordinate thus represents the number of items with a measurement *equal to or less than* the corresponding abscissa. Then Galton connects the upper ends of the ordinates and a broken line originates on which a definite point corresponds to every magnitude occurring on the axis of abscissas. Finally, he bisects the last ordinate and draws a horizontal line through such point parallel to the base. The abscissa of the point of intersection of the horizontal and the broken line is the median. In a similar way the quartiles and deciles of the series may be found.

By joining the tops of the ordinates with *straight lines* it is assumed that the distribution of the items within the classes is uniform, and the values of the quartiles and deciles found by such graphic interpolation depend upon that assumption. We may also connect the tops of ordinates by as regular a curve as possible, or draw a curve which (even if it does not contain all the points, nevertheless passes very close to them) fits the configuration of the broken line as closely as possible and clearly represents the formation of the series intimated in the diagram. From such an adjusted curve the median may be determined as explained in the preceding paragraph.

Galton's method of the graphic determination of the median is the counterpart of his "cumulative" groups.

These groups are obtained by successively summing the frequencies as given in the usual table, starting with the lowest or highest class.[90] In the graphic method of Galton, it is true, no numerical sums were ascertained arithmetically, but as every ordinate commences at the elevation of the end of the preceding ordinate, the total ordinates (measured from the basis of the diagram) show how many items belong to every given class together with those belonging to all the preceding classes. It is a graphic representation of the numerical sums which would be found if, starting from the lowest class, the single classes were added successively.

Bowley has applied Galton's method to the determination of the median from a series of wages.[91] In the diagram constructed by Bowley, the abscissas correspond to the different wages under consideration, the ordinates do not correspond to the numbers of workmen with wages of a certain size, but to the numbers of workmen earning at or above the wage represented by the abscissa. The line joining the tops of the ordinates has its lowest point in the highest wage class at the right of the diagram, where there are only a few workmen who earn that wage or more. The line ascends continuously toward the left end; for the lower the wages the greater is the number of workmen earning at or above this wage. Since the height of the diagram corresponds to the total number of items, therefore, by bisecting this height, we can immediately ascertain that point of the diagram the abscissa of which represents the median.

Bowley has also adjusted the diagram just described. The median determined from the unadjusted diagram was $1.49, the same as the median computed by the elementary arithmetic method from the original series of numbers.

[90] See details on this method of formation of groups, p. 85 f.

[91] Elements of Statistics, 2nd ed., p. 154 f. The series which he uses represents the wages of 5,123 American workmen taken from U. S. Senate Report on Wages, Prices, etc. (1893).

From the adjusted curve the median \$1.51 was obtained; mathematical interpolation gave \$1.536.

4. APPLICATION OF THE MEDIAN

The application of the median to represent statistical series is far less extensive than that of the arithmetic mean. However, it is used in various fields of statistics to represent important quantitative phenomena.

In the field of population statistics the median frequently serves to represent the age constitution of various masses of population. The most important application of the median in this field is the "probable lifetime" (wahrscheinliche Lebensdauer, vie probable or vie médiane, vita probabile or vita mediana), which is the median computed from the length of lives given in the mortality table for a definite generation. The probable lifetime tells what age half the individuals of a generation, born contemporaneously, will survive. It may be ascertained for the new-born as well as for persons in other age classes. Since the mortality tables usually contain age classes of one year each, the principles indicated above for the determination of the median in a series consisting of classes must be observed in the computation of the probable lifetime.

The probable lifetime of a population is usually longer than the expectation of life. The reason for this is that the very high death rates in the age classes above the probable lifetime remain without influence upon the probable lifetime (the median), while they must essentially reduce the expectation of life (the arithmetic mean).[92] The probable lifetime does not express at all the mortality in the older age classes, since, according to the nature of the median, the numerical values of the items located away from the center of the series have no influence upon the magnitude of the median. Therefore, the probable life-

[92] Cf. v. Mayr, Bevölkerungsstatistik, p. 268.

time may remain the same even if the mortality of the older age classes changes considerably. The expectation of life, on the contrary, being the arithmetic mean of the lifetimes under consideration, depends on the values of all the items contained in the series; therefore, it is influenced by the mortality in the highest classes and reflects every change in them as well as changes in other classes.

Of special influence on the probable lifetime is the intensity of the mortality of children. If the mortality of children in the first year—as it happens in some districts—is 50% of those born, then the probable lifetime of the newborn is one year.[93] The expectation of life is, of course, much higher, since it depends also on the numerical values of the lifetimes of those persons not having died in childhood. In contrast to the intensity of the mortality of children, the distribution of this mortality over the different age classes has no influence upon the probable lifetime, of course, on the condition that the probable lifetime itself is not located in the younger age classes. Therefore, the probable lifetime expresses neither changes in the mortality of the young age classes nor of the older age classes.

Sometimes the median is also computed from the ages given in death registers. In former times this median frequently was called the probable lifetime, just as the arithmetic mean of the ages of the dead was called the expectation of life. However, it is now an established fact that the mean and the probable duration of life can be computed correctly only from mortality tables.

It is self-evident that the median may also be computed from the age constitution of those living. Lexis calls this median the " central age " and thinks that this value is just as well adapted for the general characterization of the age conditions of the population of a country as the arithmetic mean age of those living, which can be ascertained only after tedious arithmetic operations. Further-

[93] Ibid. p. 267.

more, the chances are even that the age of an individual chosen at random from the population exceeds (or is less than) the central age. Therefore, the central age may be called the probable age of those living contemporaneously.[94]

Likewise, the median of the ages of those marrying is of interest. It may be called the probable marrying age. Boeckh has also determined the probable *duration* of marriage (the median) from his frequency table of marriages in Berlin according to their duration.

The application of the median has recently been extended to the field of wage statistics. Fox was the first to use the median of wages extensively in his *Report on the Wages and Earnings of Agricultural Laborers in the United Kingdom* (1900). In that report (p. 25) the median wage rate is defined as the "rate so chosen, that the numbers of laborers, whose rates of wages are above and below that rate, respectively, are, as nearly as possible, equal." In the *Journal of the Royal Statistical Society*[95] an anonymous reviewer of the report (Bowley?) welcomed the use of the median with the following words: "All Statisticians will be glad to see this most useful average at last boldly and explicitly used in an official publication." Also in the *Second Report by Mr. Wilson Fox on the Wages, Earnings, and Conditions of Employment of Agricultural Laborers in the United Kingdom* (1905) median wage rates are presented in a similar way as in the first report.[96]

[94] Lexis, Zur Theorie der Massenerscheinungen, p. 36.

[95] Vol. LXIII (1900), p 505.

[96] In the first report the median rate was also called the predominant rate. To this name the reviewer in the Journal of the Roy. Stat. Soc. had justly objected, since by this term we denote the most frequent wage (the mode in the series of wage data). But, obviously, the relatively most frequent and the median wage need not coincide. In the second report the term predominant rate for the median wage is omitted. However, it is stated that the median rate "in almost every country corresponded to the predominant rate,

The median is also used extensively in the wage statistics of the United States Bureau of the Census of the year 1903,[97] and that principally (in Chapter II) to compare the wage conditions in the years 1890 and 1900. The quartiles of the series in question are used here as complements of the medians. When computing the median the problem was to find the " employee who stands halfway between the lowest-paid and the highest-paid employee."[98] The wage-class to which this employee, standing in the center of the series, belongs, is easily found. However, in the statistics quoted the particular wage of the halfway individual is not ascertained within the limits of the wage class under consideration, but the inferior limit of that wage class is taken as median. Although this is a simplified procedure it is not theoretically correct.[99]

It is possible that the median will be more widely used in the field of wage statistics in the future. Bowley favors the use of the median and illustrates the methods of the determination of the median usually by series of wage statistics.[100]

that is the rate at which the largest number of laborers in each county was paid" (p. 27 and p. 148). It is not clear in what way the median wage rates in the two reports were computed in those cases where the median served to characterize the wages of a whole county, the computation being based upon wage data for various rural districts. In these cases individual wage data were not given, but merely "rates of weekly cash wages most generally paid to ordinary agricultural laborers" for the single rural districts, into which counties are divided. Thus, series of averages (modes) were given, from which correct medians cannot be computed for reasons which have been explained in another place (p. 204 f.).

[97] Twelfth census of the United States, special report, Employees and Wages.

[98] Ibid. p. xxvi.

[99] Ibid. p. xxxi.

[100] Prof. Mandello, as he announced in the Bulletin de l'Institut intern. de Statistique, Vol. XIII, No. 1, p. 401, has used the median in the field of historical wage statistics, published in the Hungarian language.

Of the other fields of statistics where the median is used, that of anthropological statistics is the most important. If it is desired to quickly characterize a series of anthropometric data by a mean, the median is most available. Since anthropometric series frequently have a regular, symmetric structure in which the arithmetic mean, the median, and the mode lie close together, every one of these values has a typical character. The median, however, has the special advantage that it can be determined the easiest and the quickest, i. e., by mere counting.

Nor in the field of price statistics is the median altogether unknown. It has been used in the computation of mean index numbers, i. e., in the computation of a mean of the special indices representing the price fluctuations of different commodities in the course of years.[101-102]

Finally, we may refer to the importance of the median in the theory of error, both in its application to actual errors of observation and to statistical data. In the theory

[101] Bowley recommends the use of the median in the computation of total index numbers, especially on account of its independence of extreme values which may originate from abnormal price fluctuations of single commodities (Elements of Statistics, 2nd ed., p. 224).

[102] See also W. C. Mitchell, Gold, Prices and Wages under the Greenback Standard, for further use of the median.

Irving Fisher has adopted the median in his study, The Purchasing Power of Money. He agrees with Edgeworth's conclusion that "in the present state of our knowledge, and for the purposes on hand, the median is the proper formula" (Edgeworth, "First Report on Monetary Standard," Report of the British Association for the Advancement of Science, 1887, p. 191). After making various tests of the reliability of the median Prof. Fisher says, "The final *practical* conclusion, therefore, is that the weighted median serves the purposes of a practical barometer of prices, and also of quantities, as well as, if not better than, formulæ theoretically superior" (The Purchasing Power of Money, p. 427). Edgeworth gives many arguments for the use of the median in the reference cited above. However, Francis Galton was the real popularizer of the median (see Natural Inheritance, p. 47).—TRANSLATOR.

of error the median is determined for the series of the deviations of the items from the mean and with other averages of these deviations serves as a measure of the dispersion of the series. The median of the deviations of the items from the mean is called "probable deviation" (or "probable error") and it is that deviation which is exceeded just as often as it is not exceeded,[102a] so that there is the same probability that an item chosen at random has a smaller deviation, as there is that it has a greater deviation than has the median.[103]

A peculiarity of the median, to which Galton first called attention,[104] is found in the fact that it can also be computed if different size—or intensity—scales of a non-measurable phenomenon are given. Of such series no other average but the median can be computed. An illustration follows.

Let the problem be to ascertain the average intelligence of a group of students. The different degrees of intelligence cannot be given numerically. But often it is not difficult to arrange the students according to the degree of their intelligence. Then the student standing in the center of the series represents the median intelligence of that group of students.[105]

In this way it would also be possible to compare the mental qualities of various groups of students who differ from each other in sex, social standing, descent, etc. Something definite might be established by this method, espe-

[102a] Lexis, Zur Theorie, etc., p. 24.

[103] Since one-quarter of the observations lie on each side of the mean, i. e., within the limits of the probable error, Yule and Bowley also call the latter the quartile deviation.

[104] Cf. Natural Inheritance, p. 47, and "Statistics by Inter-comparison," Philosophical Magazine, Vol. XXXIX (January, 1875), p. 33.

[105] Cf. Bowley, Elements of Statistics, 2nd ed., p. 126; Lexis, Zur Theorie, etc., p. 39 f.; Lexis, Abhandlungen, etc., VI, "The Typical Values and the Law of Error," p. 126.

cially about the relative intelligence of the two sexes.[106]

The median can also be computed in series which do not give the size or intensity of a phenomenon numerically but by descriptive terms. Bowley gives an example of the application of the median (with simultaneous utilization of the quartiles) in the representation of such a series of non-numerical data.[107]

The point in question is to characterize by an average the information about the extent of working overtime, the information being given by 88 branches of the Amalgamated Society of Engineers of 20,666 members. The data furnished by the branches only rarely give the numerical amount of overtime. Most of the answers are without numerical precision, such as "very little," "when necessary," "moderate," "rather general," etc. Bowley arranges these answers according to the extent of overtime indicated by them and selects the answer located in the center of the series. The extent of overtime given by this answer may, in a certain sense, be considered to be the average. When determining this central item, account had to be taken of the fact that the individual answers refer to varying numbers of workmen, since the individual branch societies have varying membership. The series corresponds to a series of wage data in which varying numbers of workmen belong to the different wage classes established. Taking into consideration the varying membership Bowley found the statement "maximum 18 hours in 4 weeks" or "moderately" to be the median for the characterization of the extent of overtime work; the lower quartile was "very little," the upper "14 hours when busy." Without consideration of the varying membership the median "not much," and the quartiles "very little" and "when necessary" or "occasionally" were found.

[106] Cf. Lexis, "Anthropologie und Anthropometrie" in the Handw. d. Staatsw., 2nd ed., p. 393 f.

[107] Elements of Statistics, 2nd ed., p. 136 ff.

CHAPTER V

THE MODE

1. CONCEPT AND PROPERTIES OF THE MODE

The mode, the predominant, most usual, or normal value, the mean of density, or the place of greatest density (German: "der dichteste Wert," French: usually "valeur normale" or simply "normale," also "modus"), is the value occurring most frequently in a series of items and around which the other items are distributed most densely. Fechner defines the mode as that value " around which the items and, consequently, the deviations collect most densely, so that equal intervals contain more items the nearer the intervals lie to this value, no matter if they are taken on the positive or the negative side."[108] Therefore, the mode represents the most probable value of the element of observation represented in the series. " The mode lies at a place of concentration in the series arranged according to the sizes of the items, so that the probability that an item chosen at random will belong to a group of values which contains the mode is greater than that it will belong to any other, equally large, group of items" (Czuber).[109]

If a series of individual observations is represented graphically, as usual, by a system of coordinates so that the abscissas of the single points of the diagram correspond to the various magnitudes occurring in the series, and the

[108] "Über den Ausgangswert," p. 11.

[109] Wahrscheinlichkeitsrechnung, p. 334. See also Lexis, Zur Theorie, etc., p. 27; also Fechner, Kollektivmasslehre, p. 171.

ordinates represent the frequencies of the different magnitudes, then the mode of the series is the abscissa of the maximum ordinate of the diagram. On account of this property the mode is sometimes called the "maximum ordinate average."[109a]

The mode has many properties in common with the median. It is, just as the latter, "une moyenne de position," i. e., its size, unlike that of the arithmetic and geometric means, does not depend on the sizes of all the items, but, like the median, merely on the sizes at a definite place of the series. Just as the median is given by the item in the center of the series, so the mode is given by the size of that item which, on account of its relatively greatest frequency, is considered characteristic of the whole series. The numerical sizes of the items that are distant from the place of concentration containing the mode are disregarded and the mode gives no information about the sizes of these items. Therefore, series in which the places of concentration coincide give the same modes even if the parts of the series above and below the places of concentration are entirely different in the two series. Changes that occur in the course of time in the parts of series aside from the place of concentration do not influence the size of the mode.

Just as the median, so, for similar reasons, the mode can only be computed for series of quantitative individual observations, such as series of wages, incomes, ages, etc. (series of the first of the three groups). It is only in series of this kind that the items are arranged according to value, which arrangement is required in determining the mode. In the series of the third group, consisting of relative numbers or means, the mode cannot be computed because the items of these series usually refer to varying numbers of units and, therefore, are of different weights,

[109a] Venn used this term (see article on "Average" in Palgrave's Dict. Pol. Econ.), but his usage has not been adopted.—TRANSLATOR.

these weights, however, not being ascertainable from the series itself. Therefore, the actual distribution of the various intensities and sizes cannot be found from the series, and the intensity or the size actually occurring most frequently cannot be ascertained. If, for example, a series of death rates referring to different geographical districts is given, then the rate most frequently occurring in the series may actually pertain to a smaller number of people than some other rate occurring less frequently in the series. That will be the case if the former rates are based on smaller, and the latter rates on larger, geographical districts.[110]

As far as series of averages are concerned, the sizes of the individual observations, on which the averages forming the series are based, cannot be ascertained from them. But

[110] Colajanni is one of the few statisticians who try to determine the mode also in series of relative numbers of the kind mentioned above (see Manuale di Statistica teorica, Napoli, 1904, p. 182 ff.). He distinguishes between the "ordinata massima" and the "media di densità." By the first term Colajanni means the real mode of a series of individual observations (for instance, measurements, wages). By "media di densità," however, he means some sort of a mode of a series of relative numbers. Colajanni gives an example of the computation of a "media di densità" from classes of equal breadth in a series of marriage rates for the years 1872-1879, arranged according to size. He obtained his result by computing the arithmetic mean of the class of greatest frequency. The author, however, raises the objection that series of relative numbers, for instance, marriage rates, do not admit the computation of a correct mode for the reasons just mentioned, but especially on account of the different weights of the items. The Italian marriage rates which Colajanni uses are of different weights, because the Italian population has changed numerically in the course of years. The marriage rates of the later years have greater weight on account of the growth of population. The degree of intensity occurring most frequently in the series, if the values in question refer to the earlier years, may be of less importance, for the purpose of computing an average, than a degree of intensity occurring less frequently in the series, the values of which are based on the larger population of the later period.

the sizes of these individual observations are necessary if we want to determine the most frequent individual item. If, for instance, a series of average—or normal—wages for the different districts of a country is given, then it is impossible to determine from these values the relatively most frequent, densest, normal wage for the whole country, i. e., that wage which relatively most individuals earn. It is true we can arrange the average or normal wages referring to the various districts according to their sizes and see if this series contains a place of concentration. But the mode eventually found in this way does not need to coincide at all with that mode which could be determined from the unknown series of individual wages of the workmen of the whole country and which alone could be considered to be the actual " densest " wage for the whole country.

Finally, we must recognize that, owing to its nature, the mode is significant only in series that consist of a large number of items and, therefore, can contain a " place of concentration " in the true sense of the phrase. This condition is true only in series of individual observations (series of the first group); the series of the second and third groups usually consist of but relatively few items.

The mode is not a power mean in Fechner's sense. It was pointed out in defining the mode that this mean represents the most probable value. The fact that this mean, in contrast to the arithmetic mean and to the median, can never be a mere arithmetic abstraction and always possesses more or less typical value, is, however, of greater practical importance. Owing to its conception, the mode always represents a quantity which occurs with relatively the greatest frequency in the series of items and around which the other items are grouped with more or less regularity. It is, of course, useful to know what wage relatively most workmen earn, at what age relatively most people die, marry, etc. Changes of the mode in course

of time, in contrast to changes of the arithmetic mean or the median, always mean changes in which a greater number of units of observation take part. If the arithmetic mean of a series of wage data is computed, then the disappearance of a few extreme, high or low, wage items may cause considerable change in the average. But a change of the modal wage, i. e., of that wage which relatively most wage-earners receive, cannot occur, unless the wages of numerous individuals have changed. However, if the mode has increased or decreased a certain amount, it does not follow that the same individuals are necessarily in both modal groups, and that they have all experienced personally the observed change of the densest wage.

But the mode is not merely of interest with reference to the items which it characterizes directly. The importance of the mode lies in the fact that it is the average best suited to represent the "normal" or "typical" size of a variable phenomenon. In the sense of mathematical statistics, it is true, we can speak of a typical average of individual observations only if the items show a grouping around the average which corresponds to the hypothesis of merely accidental disturbances of a normal value.[111] However, such typical means, strictly speaking, occur very seldom. The practical statistician must take into account the fact that most statistical series do not show a regular structure corresponding to the law of chance and, consequently, have no typical means in the strict mathematical sense.

Of the averages computed from statistical series, the arithmetic mean and the median very often are not typical. Frequently, they are values which do not at all, or only very rarely, appear in the series. However, the mode lies at a place of concentration around which the series is distributed in both directions with some regularity. It may be assumed that it comes nearest to that value which the general

[111] See Lexis, Zur Theorie der Massenerscheinungen, p. 38.

complex of causes would produce if free from disturbing influences. In non-mathematical statistical terminology the mode is, therefore, frequently called " typical " or " normal " value even when the distribution does not agree with the law of chance. Rümelin, in his *Bevolkerungslehre,* called the relatively most frequent age of marriage the " normal " age of marriage. In English and French works it is quite usual to call the relatively most frequent wage rate the " normal rate " or the " salaire normal." In French writings the mode is called by some authors simply " valeur normale," or briefly " normale." This statistical usage agrees with the common usage. If non-statisticians speak of a normal wage, normal income, normal price, etc., they undoubtedly think of the relatively most frequent wage, income or price. The same average is in mind when we speak of the size of an agricultural state or establishment as typical for a certain district.

Special importance must be attributed to the mode as a starting point for measuring the dispersion of statistical series. Only in series distributed symmetrically around the arithmetic mean—and in this case arithmetic mean and mode coincide—may the arithmetic mean be used as a basis for measuring the dispersion of these series. However, in series in which the mode does not coincide with the arithmetic mean we may also start from the mode in order to examine the formation of the parts of the series on both sides of it. If in such a case we start from the arithmetic mean, then we have a place of concentration on one side of it, or, in graphic representation, a " hump " of the curve, to which no similar formation corresponds on the other side of the arithmetic mean. However, if we start from the mode we frequently find a certain regularity of the series. The most pronounced case of such regularity is given when the series coincides with a skew curve of error or corresponds to the asymmetric Gaussian law, which is the case if the items are distributed on both sides of the

mode according to the normal law of error but with different degrees of precision or standard deviation.

The above-mentioned properties of the mode may, in individual cases, appear in varying degrees. While there are series which possess a prominent point of concentration noticeable at the first glance, around which the series is distributed with the greatest regularity, on the other hand we sometimes come across series which show only indistinct, hardly ascertainable, points of concentration and no regular grouping around these at all. Therefore, the scientific value of the mode, its applicability for various scientific and practical purposes, depends largely on the formation of the series in the individual case.

The series possessing no point of concentration at all, as well as the series showing two and more points of concentration, deserve special mention.

Series in which the items are distributed within definite limits without any point of concentration whatsoever, occur very rarely. Such series are found, however, in the field of anthropology. Alphonse Bertillon has based his anthropometric method for the recognition of criminals on this fact. If certain parts of the bodies of a great number of people are measured, then series of measurements originate which show different formations according to the parts of the body measured. The measurements of certain parts of the body, as experience shows, group themselves around a relatively most frequent, normal measurement (a typical mean). These measurements are not fit for the identification of individuals, because many people will show the same value, i. e., the most frequent measurement, or measurements approximating this. Therefore, the stature, for instance, possesses only small "signalizing" value and plays no important rôle in Bertillon's method of identification. Other parts of the body, however, have no typical value, and there is much less probability that the same meas-

urement occurs with many different people. Objects of measurement which produce series without typical mean are, for instance, the inside length of the leg, the width of the hips, the length of the head, the reach of the arms, the length of the foot, and the length of the middle finger.[112]

As far as demographic observations in the precise sense of the term are concerned, it seems that the group of premature deaths, distinguished by Lexis in the mortality table, which lies between the group of childhood mortality and normal old-age mortality, does not possess a place of concentration.[113] To be sure, premature deaths cannot be reproduced in an independent series, since the deaths cannot be individually assigned to the three groups of mortality mentioned. However, we can gain an idea of the number and distribution of the premature deaths, by producing the curves of childhood mortality and old-age mortality towards the center of the curve which represents the mortality table, and thus deducting a number of deaths, corresponding to the extreme end of the childhood mortality and the beginning of the old-age mortality, from the totality of the deaths of the middle-age classes. The remaining premature deaths are distributed rather uniformly over the middle-age classes under consideration without a noticeable point of concentration.

Series having two or more modes, and consequently resulting in curves with two or more culmination points

[112] With a height of from 1 60 to 1.65 meters the inside length of the leg varies from 730 to 825 mm. and the curve of distribution resulting from measurements of a great number of people is very long and irregular (Lexis, Abhandlungen, VI, "The Typical Values and the Law of Error," p. 125; see also A. Bertillon, Das anthropometrische Signalement, Lehrbuch der Identification, German by v. Sury [1895]).

[113] See Lexis, Zur Theorie, etc., p. 43, and Abhandlungen, V, "On the Causes of the Small Variability of Statistical Relative Numbers," p. 88 f.

when reproduced graphically, occur more frequently than series without any mode. The modes sometimes stand equally prominent, sometimes one mode is most prominent, but other modes of smaller importance (secondary modes) may be established. The presence of several modes in a series usually indicates that the series consists of heterogeneous constituents, the typical values of which are different from each other.[113a]

The stature of the population of some districts furnishes the best known example of a series with two modes. As early as 1863 Adolphe Bertillon proved the occurrence of several modes in the distribution of heights of the recruits of certain French Departments [113b] and the demonstration was substantiated by Jacques Bertillon in 1885,[113c] and since then investigations in other countries have led to similar results. It is supposed that this kind of distribution of heights can usually be explained by the mixture of several races of different typical heights. J. Bertillon based the strange distribution of height, showing two modes, of the population of different Departments of Northeastern France on the fact that the population originated from a mixture of Celts of a smaller typical height and Burgundians of a larger typical height. Quetelet in the 21st of his *Letters on Probabilities* (Lettres sur les probabilités) had already pointed to the fact that the distribution of heights of a population originating from the mixture of two races of different heights would indicate the fact of

[113a] Sometimes several modes are caused in homogeneous series of measurements because of concentration of the items at *customary* values. For instance, in the wage data of the Dewey special report on Employees and Wages (census of 1900) there are modes at the round number wage points, e. g., $2.00 per day or $2.50 per day. Although the cause of such modes is evident, i. e., accounting convenience, yet there is no reason why they should not be called real modes.—TRANSLATOR.

[113b] Bulletin de la Société d'Anthropologie.

[113c] La Taille de l'homme en France.

the mixture of races, although he then had no material of observation at hand.[114]

Series with several modes, however, are found not only in the field of anthropometry. Series referring to social phenomena often show several modes, especially when the totalities represented consist of several heterogeneous constituents. Thus, the duration of life given in the mortality table shows two modes, of which the first corresponds to the excessively great mortality of infants, the second to the " normal " mortality of old people. Lexis, evidently, was induced by the presence of these two modes to distinguish three groups in the totality of deaths, first, the group of childhood mortality, second, the group of old-age mortality (the normal group of deaths), both based on pronounced modes (points of concentration), third, the group of premature deaths, which, as has been mentioned above, are distributed without pronounced mode over the middle-age classes between childhood and old age.

Other series of age data, besides the mortality table, also often have two or more modes which point to the presence of heterogeneous constituents within the totality in question. Thus, the age constitution of persons engaged in certain occupations sometimes shows two modes—for instance, one in an adolescent class, the other in a much higher age class. Such a formation of the series points to the fact that principally youthful persons, whose capacity for work is not fully developed, and persons with decreasing capacity for work devote themselves to that occupation, while strong and able men keep away. Also, the age con-

[114] Not only different nations are of different types as to size, but also city and country people frequently have different heights, and also the height for different occupations is different. Therefore, the constitution of a population of city and country people and of different occupations may result in the occurrence of several culmination points. (See Dr. Siegfried Rosenfeld, "Einige Ergebnisse aus den Schweizer Rekrutenuntersuchungen," Allg. Stat. Archiv, Vol. V, p. 124.)

stitution of persons convicted of certain crimes (as arson, rape, poisoning) and the age curve of certain groups of suicides sometimes have several modes. Quetelet noticed the fact that principally youths and old men are guilty of the crime of rape.[115]

The establishment of several points of concentration in a series may, under certain circumstances, induce us to form " natural " groups which correspond to the various more homogeneous constituents contained in the total series, rather than to divide the series into classes of equal breadth. A series showing several points of concentration ought, in such cases, to be divided into " natural " groups in such a way that the latter coincide as much as possible with the constituents distributed around the various points.

The establishment of several points of concentration in a series may also induce us to separate them and to present the heterogeneous constituents of the series independently. But this, of course, presupposes that the criterion which is the basis of the distinction between the constituents was included in the observation and that every individual case, when observed, was examined with reference to this criterion. If, for instance, the wages of all the workmen of a certain district are ascertained and represented jointly, then, frequently, a series with several points of concentration is found. If two such points appear, then the constituents which are distributed around them may be distinguished with reference to the sex of the laborers. Women usually earn lower wages than men; therefore the most frequent wage for women does not coincide with the most frequent wage for men. The former lies in a lower wage class, a fact which causes two points of concentration. If the sex of the various laborers was ascertained in the investigation, then it is easy to separate men and women, i. e., to form constituent series which are homogeneous with

[115] Über den Menschen, German edition by Dr. V. A. Riecke, 1838, p. 547; see also Öttingen, Die Moralstatistik, 3rd ed., p. 519.

reference to the criterion of sex. If the presence of two points of concentration actually was the consequence of combining the sexes, then the two new series formed will show only one mode each and a more regular formation in general.[116] Wage data for workmen of only one sex may also have several points of concentration—for instance, if workmen of different occupations or categories of work with different typical wages were combined. If during the investigation the occupations of the individual workmen or the different categories of work were ascertained, then it is possible to divide the totality of the workmen into more homogeneous constituents. The constituent series formed in this way probably will have only one mode each, around which the series is distributed with a certain regularity.[117]

2. DETERMINATION OF THE MODE

The determination of the mode of a series of individual observations is very simple if the series shows a pronounced point of concentration with regular distribution of the items on both sides of it. No computation at all is required; it suffices to glance over the series to ascertain the mode. In graphic reproduction of the series a single, clear culmination point is seen, the abscissa of which can easily be read.

[116] Mortality tables which have been computed for both sexes, jointly, also show two points of concentration of old-age mortality; if the two sexes are separated, more regular curves are found with only one mode each. (See Bulletin de l'Institut international de Statistique, Vol. X, No. 1, "Vovimento della populazione in alcuni stati d'Europa e d'America," Parte II, Statistica delle morti negli anni 1874-1894. Tavole di sopravivenza, Tav. I f., III f.).

[117] Numerous illustrations of series of wage data which show several points of concentration are to be found in Bowley's Elements of Statistics. Usually (see diagram to p. 145) there are a central mode and two or three secondary modes, which correspond to groups of workmen with wages either far above or far below the average.

However, the series originating from statistical observations usually are not of an entirely regular structure, and, unless the statistician is working on the original material of a definite statistical investigation, he, in most cases, does not have to do with individual observations given as such. On the contrary, he has to work with series that consist of classes and merely show the number of observations belonging to the different classes. Now this fact is not at all unfavorable to the determination of the mode, as we might, perhaps, think at first. As has been mentioned, the individual items of a statistical series hardly ever show an entirely regular structure, and, in the compass of the relatively most frequent items, very irregular fluctuations are found which render it impossible to determine the *exact* mode on the basis of the individual items. Only by the formation of classes, in which the accidental fluctuations counterbalance each other, does the series receive a more regular structure which reveals the real points of concentration. Therefore, if a series, the mode of which is to be determined, does not consist of classes from the beginning, but shows the original individual data, then it is expedient in most cases to first condense the series in well chosen classes. To be sure, the result of this procedure is the further task, sometimes connected with difficulties, of determining the exact position of the mode in the class of the greatest frequency.

A strange statistical phenomenon is found in the fact, observed especially by Bowley, that the position of the mode is dependent on the nature (breadth and position) of the classes of which the series consists. If we condense the same original material several times into classes of different breadths, or if we form classes of the same breadth but of different limits, then it may happen that the series thus formed have different points of concentration. Therefore, in order to ascertain the correct mode, it is sometimes expedient to condense the same items repeatedly into

classes of varying breadths and positions in order to find by experimenting (méthode de tâtonnement, Liesse) that kind of group formation which is best suited to the structure of the totality and, consequently, represents the position of the mode most correctly. The arithmetic mean and the median can be computed the more accurately, the more detailed are the classes in which the items are distributed. However, this does not hold for the computation of the mode. Classes of greater breadths may show the mode more plainly than more detailed groups of lesser breadths. In illustration of this Bowley gives in his *Elements of Statistics* (p. 119) the following example, which, it is true, is based on a very irregular series, so that the computation of the mode is connected with special difficulties.

The point in question is to ascertain the mode of a series taken from the U. S. *Report on Wholesale Prices, Wages and Transportation* (1893). Bowley tabulates the wages of 5,123 workmen.[118] He combines the items successively into classes of 10 cents, 20 cents, 30 cents, and 50 cents in breadth.[119] In the series of 10-cent grouping the wage class \$1.15-\$1.24 has the greatest frequency; to it belong 685 workmen. However, the series of 10-cent groupings is, as Bowley explains, of very irregular structure; the series contains, strictly speaking, 14 points of concentration, of which, however, the one in the wage class \$1.15-\$1.24 is the most prominent. In the series of 20-cent classes the numbers of workmen belonging to the different classes, starting the lowest class at 25 cents, are: 16, 144, 270, 370, 989, 557, 538, 531, etc. The number 989 for the wage class \$1.05-\$1.24 means a pronounced mode. If we start the lowest class at 35 cents, we obtain the numbers: 74, 242, 282, 505, 784, 924, 274, etc. The mode lies in the wage class \$1.35-

[118] This is the same series of wage data which Bowley has used also in the graphical determination of the median. (See above, p. 214).

[119] See the tables (p. 91 and p. 120) in Elements of Statistics.

$1.54, to which 924 workmen belong. Thus two series, both of which contain classes of 20 cents breadth each, show the mode in quite different wage classes. This shows that the mode cannot be ascertained at all by forming classes of this breadth. If 30-cent classes are formed, then we obtain, according as we start from 55 cents, 65 cents, or 75 cents when forming the groups, the following values: 355, 674, 1,242 (for the wage class $1.15-$1.44), 740, etc., or 439, 1,190 (for the wage class $0.95-$1.24), 1,023, etc., or 483, 1,088 (for the wage class $1.05-$1.34), 996, etc. The wage classes of the greatest frequencies in the three series are: $1.15-$1.44; $0.95-$1.24; and $1.05-$1.34. The three wage classes partly coincide, inasmuch as all three contain the narrower wage class $1.15-$1.24. Therefore, we may assume that this wage class contains the mode. It would lie, perhaps, in the middle of this wage class, and thus would be about $1.20.

Bowley summarizes the method of computing the mode described above as follows: "Tabulate the figures again and again in gradually widening groups till regularity is obtained; then examine again the groups which have the selected width and see if the mode is shifted when the lower limit of the grouping is moved; if it is shifted the groups are not wide enough; if it is not, the mode is in the smallest group common to the larger equal groups which all contain it."[120] Of course only statistical bureaus having the original material at hand are able to follow Bowley's method. These only are able to form experimental classes of any breadth and position. The statistician who works on a series contained in a statistical publication can form groups of greater breadth merely by adding adjoining classes, but he cannot divide the given classes nor change their limits except by means of hypothetical interpolation.

Having found the class in which the mode of the items

[120] Ibid. p. 121.

lies, we can determine the mode more accurately within the limits of this class.[121] In the illustration quoted above (see p. 236) Bowley has assumed that the mode lies in the center of the class of the greatest frequency. This assumption corresponds to the hypothesis that the items within the limits of the class of the greatest frequency are symmetrically distributed around the mode. But the structure of the series may be such that the items may have an asymmetrical distribution within the limits of the class of the greatest frequency. Therefore, in such a case, another hypothesis, more suited to the structure of the series, may be chosen. We may also resort to graphic, or, if we want to proceed with very great accuracy, to algebraic interpolation. Bowley has found the mode in the series of wages of the American workmen, by means of algebraic interpolation, to be $1.10,[122] while the elementary method gave $1.20.

As is known, statistical series are frequently subjected to adjustment in order to make their characteristic formations appear more clearly and to remove the accidental irregularities in their structures. The adjustment may affect the numerical value of the mode. The mode of the adjusted series or (in case of graphic adjustment) of the adjusted curve may be different from the mode of the unadjusted series or the unadjusted broken line. Since the series receives a more regular formation by adjustment, the mode of an adjusted series is more clearly seen and can be computed with greater accuracy. Since adjustment removes accidental irregularities, the mode of an adjusted series possesses greater theoretical value than the mode ascertained from the original material of the series. It approaches that ideal mode which we would obtain from an infinitely large number of observations with infinitesimal

[121] In tabular presentation of series the class of greatest frequency is usually set off by heavy print or print of different color.

[122] Elements of Statistics, 2nd ed., p. 254.

differences between the individual measurements. If the adjustment is done by means of an analytical formula, then the mode may also be computed directly from it as the maximum of the function used.[123]

A peculiar graphic method of computing the mode is based on that kind of group formation, advocated especially by Galton, in which we are not shown how many cases belong to each class, but what numerical sums we obtain, if, starting from the lowest or highest class, we add the single classes successively and form so-called "cumulative classes."[124] Bowley explains this method of computing the mode in his *Elements of Statistics*[125] on the basis of the same American wage data which he has also used in the graphic determination of the median and which have served above (p. 235 f.) to illustrate the dependence of the mode on the kind of group formation. These wage data condensed into "cumulative" classes are reproduced graphically by Bowley so that the abscissas of the single points of the diagram correspond to definite wage grades while the ordinates indicate the numbers of workmen earning at or above a certain wage. The line obtained in this way ascends throughout from right to left, since the axis of the abscissas indicates the different wage grades from right to left in decreasing order and the number of workmen earning at or above a certain wage, naturally, is the greater, the lower the wage.

The mode is indicated in this diagram by that point at which the greatest number of workmen is added; it is

[123] See especially Fechner, Kollektivmasslehre, pp. 183-186, and Lucien March, "Quelques examples de distribution des salaires," Journal de la Société de Statistique de Paris, 1898, pp. 199 f., 205, and 247.

[124] Cf. details about this method of group formation, p. 85 f.; and about the use of such cumulative classes in the graphic determination of the median, p. 213 f.

[125] 2nd ed., p. 155.

that place where the diagram is steepest. After we have adjusted the diagram and drawn a curve instead of the broken line, the tangent crosses the curve at this place. The point of crossing can be found mechanically by placing a ruler to the curve and turning it until it crosses the curve. The mode and the secondary points of concentration can also be recognized by the fact that at the points in question the curve changes from concavity toward the axis of abscissas into convexity. Bowley's curve is concave to the base-line from $0.30 to about $1.20, convex from about $1.20 to about $3 15, then again concave till $3.40, and then convex till its right end. Thus it has two points of concentration, one at about $1.20, the second, which is of much less importance, at about $3.40. The investigation of the numerical material, as carried out above,[126] has also placed the mode at $1.20. By means of algebraic interpolation Bowley has found the two modes to be $1.10 and $3.20.[127] These divergences again show how difficult it is to determine accurately the position of the mode and of the secondary points of concentration.

Bowley advocates the above graphical method of determination because the computation of the mode from the numerical material of the series meets with difficulties on account of the unequal distribution of the items on both sides of the mode and on account of the dependence of the numerical size of the mode on the kind of group formation chosen in the individual case. He explains that, when this graphic method is used, the first of these difficulties is removed entirely and the second is decreased, since with the use of this method only unessential changes of the mode according to the kind of adjustment of the diagram can occur. Another advantage which Bowley [127a] attributes to this graphic method of computing the modes, is the following: " This method can be applied to numbers which are given at irregular intervals of graduation (e. g., 30 at 30 s.

[126] P. 236 f. [127] Elements, p. 254. [127a] Elements, p. 156.

6 d., 40 at 30 s. 8½ d., 35 at 40 s. 1 d., etc.) as easily and by exactly the same construction as to more regular returns; and if the smooth curve is regularly drawn, the number of modes can be seen at a glance and the individual importance of each can be estimated.''

3. APPLICATIONS OF THE MODE

The determination of the mode has become customary in many fields of statistics. To be sure, the mode, as has already been mentioned, is frequently not determined exactly, but only the class of the greatest frequency is emphasized.

In the field of population statistics series of age data frequently suggest the computation of the mode which indicates the '' normal age '' of the group of population in question. Thus, the mode is often computed from the age constitution of those marrying and is called the '' normal age of marriage.'' This normal age of marriage obviously varies for each of the two sexes and also for different social groups.

One of the most interesting applications of the mode is the '' normal length of life '' (called also '' normal lifetime,'' '' normal age ''), a notion which Lexis has introduced with great success into statistical literature.[128] The normal length of life is the mode of the series of lifetimes contained in the mortality table, it is that age at which—with the exception of infancy—most people die according

[128] Lexis has fully explained the term, normal length of life, in several of his writings and advocated its use in the Paris demographical congress of 1878 (see Zur Theorie, etc., 1877, pp. 42-64; Annales de démographie internationale, 1878, p. 447; 1880, p. 481, and 1881, p. 233: "Sur les moyennes appliquées aux mouvements de la population et sur la vie normale"; Abhandlungen, VI, "The Typical Values and the Law of Error," pp. 10-119. See also Czuber, Wahrscheinlichkeitsrechung, p. 337 ff.).

to the mortality table and may, therefore, be considered to be the normal, typical length of life.

The normal length of life in Germany is 71 years for both sexes combined. In most of the other countries it also lies at the beginning of the seventies. The normal length of life of women usually is considerably greater than that of men. In Germany the difference is about two years.[129]

It is especially interesting that the deaths above the normal age and those in the age classes next below the normal age—the deaths of the normal group of mortality distinguished by Lexis—are usually distributed symmetrically and according to the law of error about their mode. Therefore, the latter appears to be a typical mean in the strict mathematical sense, if not with regard to the whole mortality table, at least with regard to a considerable part of it, i. e., with regard to the group of normal mortality of old age. Indeed, the normal length of life is the only average of a typical character which can be obtained from the mortality table. Mean and probable length of life, which are based upon all the values contained in the mortality table, are non-typical, i. e., they fall in age classes which are of low frequency in the series of lifetimes. This is caused by the facts, that the totality of deaths consists of the three groups of childhood mortality, premature deaths, and normal old-age mortality, as distinguished by Lexis, and that masses which are composed

[129] For Austria two ungraduated mortality tables are given, computed according to somewhat different methods on the basis of the results of the census of December 31, 1900. (See Die Ergebnisse der Volkszählung vom 31 Dez., 1900, Österr. Statistik, Vol. XLV, No. 5, supplement.) The computation of one of these tables is based on the average mortality of the six years preceding the census; the computation of the other is based on the mortality in the two years before and after the census year. The former table of mortality shows the largest number of deaths for both sexes in the 71st year of age, the latter also for both sexes at the end of the 70th year.

of heterogeneous constituents are not regularly distributed as a whole around a " typical " mean.

Besides the age constitution of the mortality table, the age constitution of the deceased usually contains a mode corresponding to the typical old age. But just as the average age of the deceased is of lesser scientific value than expectation of life computed from the mortality table, the mode in the age constitution of the deceased is of less value than the corresponding mode in the series of lifetimes of the mortality table.

The determination of the mode is of considerable importance in the field of wage statistics. Evidently it is of great importance to know what wages the relatively greatest number of workmen earns. This wage can be considered to be the normal, typical wage. As a matter of fact the most frequent wage is usually computed in modern works on wage statistics that are based on individual wage data—for instance, in the Austrian wage statistics for the miners of the Ostrau-Karwin coal district.[130] Bowley has illustrated his methods of the computation of the mode by use of wage statistics. In his works in the field of historical wage statistics Bowley has tried especially to establish the fluctuation of the " normal wage rates " (predominant rates) of the English workmen in the course of the 19th century. Therefore he has used, as much as possible, the lists of " majority rates " published by the employers' associations, which state the wage rates on the basis of which relatively most workmen are paid in the

[130] See Arbeiterverhältnisse im Ostrau-Karwiner Steinkohlenreviere, k. k. Arbeitsstatischen Amte im Handelsministerium, 1st part, Vienna, 1904, p. 29 ff. In Table V of the publication mentioned the miners investigated are divided into wage classes according to the size of their wages, and the classes are designated to which relatively most workmen of the different categories of workmen belong. See also p. 60; there the strongest groups are given on the basis of Table IX, in which the miners are divided into groups according to their annual incomes from their work.

various occupations.[131-132-132a] In modern Italian statistics, too, the relatively most frequent wages (i salari più frequenti) of definite categories of workmen are frequently determined.[133]

Estimated modes have an extensive use in the field of wage statistics. Frequently, not series of individual wage data are procured—from which the mode can be computed afterwards—but in many instances the mode itself forms the primary object of the question, and this compels the person asked to estimate the mode. Thus, experts are frequently questioned as to the relatively most frequent, normal wage of the workmen of their district. Since they have no individual data at their disposal, they have to resort to estimates.[134]

[131] Bowley and Wood, "The Statistics of Wages in the United Kingdom during the Nineteenth Century," Pt. XIV, Journal of the Royal Statistical Society, Vol. LXIX (1906), Pt. I, p. 155.

[132] Also Prof. Mandello has used the "modus" in his works in the field of historical wage statistics (published in the Hungarian language); see the Bulletin de l'Institut international de Statistique, Vol. XIII, No. 1, p. 401.

[132a] The mode has also been used recently by the English Board of Trade in their investigations of the wages and hours of labor, rents and housing conditions, retail prices of food and the expenditure of working-class families on food in the more important industrial towns of important countries. Reports have been issued for the United Kingdom (Cd. 3864), Germany (Cd. 4032), France (Cd. 4512), Belgium (Cd. 5065), and the United States (Cd. 5609). —Translator.

[133] See "I salari e gli orari più frequenti per gli operai organizzati in Genova," Bolletino dell' Ufficio del lavoro, May, 1905, p. 735.

[134] In the investigation, already mentioned, of the conditions of workmen in the Ostrau-Karwin coal district, which the Austrian Labor Statistical Bureau made in 1901, the individual earnings of the miners were ascertained; but for purposes of comparison, data on the conditions of workmen in small industrial establishments in the same district were also obtained, by means of an interrogatory concerning the normal wage per week for workmen of various indus-

Also, individual workmen are sometimes asked about their normal wage, for instance, per week during the last year. They would be able to compute the relatively most frequent weekly wage from the " series " of the weekly wages which they have received in the course of that year, provided they have kept a record of the single weekly payments. But this is not often the case. Therefore the workmen usually will make a more or less arbitrary estimate of their " normal " weekly wage.

The relatively most frequent working time deserves the same consideration in the presentation of the conditions of working hours as does the relatively most frequent wage in the presentation of wage conditions. Figures giving the relatively most frequent working time in specified occupations are to be found, for example, in the bulletins of the Italian labor bureau.[135] Also in other publications the relatively most frequent working time is frequently taken into due consideration.[136]

In price statistical publications the price is frequently stated at which the largest number of units of a commodity are sold. Presumably this price is meant when the " usual " local price of a commodity is computed from detailed data or is asked for directly.[137] Also in the field

tries and categories for the year 1901. (See Arbeiterverhältnisse im Ostrau-Karwiner Steinkohlenreviere, 1st part, Vienna, 1904, p. li.)

[135] Cf. in the May number, 1905, p. 735, "I salari e gli orari più frequenti per gli operai organizzati in Genova."

[136] Thus in the "Statistik der Stadt Zürich," No. 1 (published by the Statistical Bureau of the City of Zürich in 1904), we learn that the net working time of the workmen in municipal service is 8½ to 12, most frequently ten hours, in summer time, and 7½ to 12½, most frequently nine hours, in winter time. (See Soziale Rundschau [Vienna], 1905, Pt. I, p. 5.)

[137] In the investigation of the Austrian Bureau of Labor Statistics in 1901 concerning the conditions of the miners in the Ostrau-Karwin coal districts there was also used an interrogatory to ascertain the local retail prices of certain commodities. The usual

of historical price statistics the mode has been used, though rarely, for the computation of total index numbers, i. e., of averages of the indices which represent the price fluctuations of certain commodities for a number of years.

Finally, we may point to the frequent use of the mode in anthropological statistics. Series of anthropometric data, as a rule, contain one or more [138] visible points of concentration which must be especially emphasized. However, there also occur anthropometric series without a mode, as has been mentioned above (on p. 228).

The mode is of special importance, since it is that average which is easiest to estimate and, therefore, can easiest be obtained in an investigation by direct questioning. It has been mentioned that in investigations the questions asked are frequently for the "predominant," "prevailing," "normal," or "usual" price or wage, and these questions are asked of persons who are considered to be especially competent to correctly estimate the "normal" wage and the "usual" price on the basis of their experience. The normal or usual magnitude is often asked for because it is much easier for an expert to estimate the mode than any other average. To determine a geometric or an arithmetic mean it is necessary to know all the individual cases, since these must be added or multiplied. Also, to determine the median it is necessary to be able to survey the sizes of the individual cases enough to find the item which lies in the center of the series arranged according to the sizes of the items. The computation of the means mentioned (arithmetic mean, geometric mean, and median) therefore presupposes that the items, the mean of which is to be computed, are known.

prices were asked for, and in case of great fluctuations of price also the lowest and highest prices. (Cf. Arbeiterverhältnisse im Ostrau-Karwiner Steinkohlenreviere, Pt. I, p. xliii.)

[138] See above on anthropometric series with two points of concentration, p. 229.

If this is the case these means can be computed correctly. If, however, the items are not known—and we are considering this case—then the means mentioned cannot be estimated at all or only with great difficulty. None of these means may have actually made its appearance in a definite individual case, and the person asked is usually not able to give any information about any of these means on the basis of his own perception, on the basis of his memory. Therefore, direct estimation would have to be quite arbitrary on account of the lack of information. In order to get at a value in an indirect way the person asked would have to estimate the sizes of all the individual cases to be taken into consideration and then compute the arithmetic or geometric mean or the median from these estimated items, a procedure which would be much too intricate for practical purposes.

The relatively most frequent value is quite different. The relatively most frequent, or "normal," size of a phenomenon impresses itself, even upon the passive observer, as a fact of experience, and every expert is able to state from memory and without further computation *that size which he has perceived most often.* Therefore, if individual observations (individual data) are not given and if the mean size of a phenomenon is to be estimated, we must ask for the relatively most frequent size in a theoretically correct way, because only this average admits of direct estimation.[139]

[139] In this sense the questionnaire planned by Marcus Rubin for demographic observations in countries without a census, and meant to be answered by travelers, missionaries, etc., contains the following questions:

À quel âge se marie-t-on [ordinairement? les hommes? les femmes?]. (Question 27.)

Peut on estimer le nombre d'enfants qu'a ordinairement une épouse? (Question 28.)

Sur les explorations démographiques à exécuter dans les pays où

However, this postulate is not always taken into account. Frequently we ask for an estimate of the arithmetic mean of definite phenomena. But the different kinds of means are not always sufficiently distinguished in practice. The persons intrusted with the estimation of an arithmetic mean often choose that mean which is easiest to compute. With estimated means, it is, therefore, often doubtful whether they actually correspond to the arithmetic mean or to the mode.

The fact that of all means the mode is the easiest to estimate makes it the most widely used mean of everyday life. If the man in the street wants to characterize a variable phenomenon by a single expression he usually resorts to the relatively most frequent size which has clung to his memory, and he feels instinctively that this value has a special importance, that it indicates, so to speak, the normal case of the phenomenon. And while doing this even less educated persons are usually conscious of the fact that the relatively most frequent case, which they are using to characterize a phenomenon, is not identical with its arithmetic mean. The author has made numerous psychological experiments concerning this point. He has asked conductors of tramways, omnibuses, etc., how long they average for covering certain distances and, frequently, has received the answer that the conductor could not say exactly, that he needed " sometimes longer, sometimes shorter, *usually* so and so many minutes." If we ask different people at what time they get up, go to sleep, or similar question, we frequently receive the answer, that they could not state the exact time when this takes place, but *usually* at such and such a time.

il n'existe pas encore de recensement. (Report of the 9th session of the Intern. Stat. Inst. in Berlin, 1903.)

PART III

DISPERSION ABOUT THE MEAN OR AVERAGE

CHAPTER I

PURPOSE AND MEANING OF ESTABLISHING THE DISPERSION OF STATISTICAL SERIES

When investigating the dispersion (grouping, distribution, spreading) of the items of a statistical series around their mean either of two different aims may be pursued. The object may be closer characterization of the mean or a closer characterization of the series.

Every average gives certain information about the series from which it has been computed, but it does not express the formation of the series. Series very differently constituted may result in means of the same numerical size. This is the reason why so many people are directly opposed to the use of means. Numerous statisticians prefer the presentation of statistical masses by the use of frequency tables wherever possible and are opposed to the use of means. However, against this tendency we can raise the objection that means are indispensable for certain purposes, especially in cases where we cannot work with entire series, and this is the reason for their frequent use. It is true, however, that averages must always be used with great caution. The question is not that of eliminating averages but of establishing the conditions and precautions concerning their use.

To these cautions belongs the investigation of the dispersion of the series around the mean used in an individual case. The value of the mean depends essentially on the dispersion of the items around it and on its position in the series. The question whether or not the mean can be considered to be "typical," can only be answered by

examination of the dispersion of the series around it. If it is found that the mean is "typical," then its use seems to have a sufficient scientific basis. In this case the kind of dispersion of the series can most easily and, indeed, most satisfactorily be characterized by a single term such as the mean or probable deviation. If no real typical mean with symmetric distribution of the items around it is present, we may yet discover a definite mathematical law of distribution, such as the skew curve of error, to which the series corresponds. Thus, the examination of the dispersion of the items enables us to determine exactly the theoretical value of the mean and its adaptation for further purposes—for instance, for purposes of comparison. It furnishes us, moreover, with a welcome supplement to the information given by the mean itself. It must, however, be remarked that the non-mathematical and the mathematical statisticians in general approach the examination of the dispersion of the items around their mean with essentially different feelings. To the non-mathematical statistician the series itself is always the most important thing. Of necessity he works with a mean in an individual case, but he treats it with a certain distrust and in different ways he tries to obtain as many data as possible about the series itself and uses these as supplementary to the means. By establishing extreme cases, sometimes also by quoting certain classes in addition to the mean, he tries to get on safer ground than the mean alone seems to offer. The mathematical statistician proceeds differently. To him the details of the series are raw material, the quintessence of which is to be expressed by a typical mean. If he succeeds in establishing such a mean or in discovering some law of distribution to which the dispersion of the items around the mean corresponds, then the mean in combination with the law of dispersion possesses independent scientific meaning and is more valuable than the most accurate reproduction of the whole series.

On the other hand, the measurement of the dispersion of a series can be made not for the purpose of supplementing the mean, but for the purpose of estimating the value of the series itself. In this case the measurement of the dispersion is a purpose in itself and the mean is only computed to obtain a suitable basis for the analysis of the formation of the series. As has already been mentioned,[1] it is often quite essential to start from the mean in obtaining a picture of the dispersion of the items. The dispersion of a series of quantitative individual observations indicates the kind and the degree of variability of an element of measurement (for instance, height, wage, etc.). The measurement of the dispersion of such series is of prime importance in the field of biology, where it leads to the statistical comprehension of the laws of variation. The dispersion of time series gives a measure of the steadiness or the variability of the phenomena in question during the course of time. The dispersion of geographical series enables us to ascertain to what variations the phenomena in question have been subject in the districts under consideration.

Important politico-economic interests are often closely connected with the degree of variability of certain phenomena and numerous measures of public administration must take this into consideration. Violent time variations in production or sale evidently must produce a shock to the economic organism which affects many people. This is especially true of wage fluctuations. For this reason the sliding wage scales, formerly in use in various industries in England and the United States, which made the wages dependent on the price fluctuations of the product, were usually successfully opposed by the workmen who suffered under the great wage fluctuations. Quetelet in his *Physik der Gesellschaft* (Physics of Society)[2] demanded special

[1] Cf. p. 125 f.

[2] German edition by Dr. V. A. Riecke (1838), p. 612.

measures to decrease the fluctuations of the price of cereals, reasoning as follows: "Since the price of cereals is one of those causes which has the greatest influence on the mortality and reproduction of man, and since this price even to-day shows the greatest fluctuations, therefore it is the duty of every far-seeing government to counteract all causes which bring about those considerable fluctuations in the price of cereals and consequently in the 'elements of the social body.'" Quetelet in the work just quoted (p. 613) finally came to the following general result: "One of the principal effects of civilization is the constant narrowing of the limits within which fluctuate the various elements upon which man is dependent. The more enlightenment spreads, the smaller become the deviations from the mean, and the nearer we approach to the beautiful and good." This general conclusion, however, which is closely connected with Quetelet's conception of the average man as the type of the beautiful and good, does not appear to be correct. In many fields we are not striving primarily for the removal of the deviations, but for progress all along the line.

The degree of constancy must be taken into consideration, especially with preliminary estimates for the future, which individuals as well as public administrations are often forced to make, for instance, in State or national budgets. If a phenomenon (such as the price of a certain commodity) has been subject only to small fluctuations in the past, then —under the supposition that no new disturbing factor appears—a conclusion as to the future is evidently more surely drawn than with phenomena which have already shown great fluctuations in the past.

The customs duties offer another illustration of the importance of the dispersion of certain phenomena in governmental administration. Many states have, as is known, largely supplanted ad valorem duties by specific duties. The ad valorem duties have been retained, however, in

some states for those goods whose value fluctuates widely.

Finally, it may be mentioned here that statistical investigations concerning the degree of constancy of so-called "moral-statistical" phenomena have led to philosophical inferences of far-reaching importance and to violent controversies, especially over the question as to whether social phenomena are ruled by inviolable laws, and the question of individual free-will.

The investigation of the dispersion of statistical series is, at any rate, necessary, inasmuch as these series show the greatest variety in this respect and as there are hardly two series of items of like distribution. Even the same phenomena, if presented statistically from different points of view, often result in entirely different dispersions. Thus, a phenomenon may show only small deviations from the average with regard to time, i. e., it may be constant, while it shows great differences for geographical or social differences. Again, other phenomena fluctuate considerably with time, but their structure remains the same from year to year.

Mathematical and non-mathematical statisticians have taken up the investigation of the dispersion of statistical series around their means. Mathematical statisticians usually examine the dispersion of the series with reference to the theory of errors of observation or the theory of probability. Numerous works dealing with these questions have been written, so that this subject forms one of the most highly developed chapters of mathematical statistics.

The following discussion of the various methodological questions of importance in the measurement of the dispersion will be based on the division of series into the three groups defined in the first part of this book.

CHAPTER II

THE DISPERSION OF SERIES OF QUANTITATIVE INDIVIDUAL OBSERVATIONS

A. MEASUREMENT AND PRESENTATION OF THE DISPERSION OF SERIES OF QUANTITATIVE INDIVIDUAL OBSERVATIONS BY MEANS OF ELEMENTARY MATHEMATICAL METHODS

The first of the three groups of series in our classification contains the series consisting of quantitative individual observations, such as measurements of age, length of life, wages, etc. The simplest way of obtaining information about the dispersion of such a series around the mean is to ascertain the extreme cases occurring in the series, i. e., the maximum and the minimum of the series as well as the average. These values, therefore, are very frequently stated in statistical publications for the purpose of concisely characterizing the dispersion of wage, price, and other series. The highest and lowest wages and prices that could be ascertained are stated as well as the average wages and prices. The highest and the lowest temperatures registered during the year, month, or day are given in addition to the mean temperature of a locality. Instead of stating maximum and minimum, or perhaps in addition to this statement, the distance between the two extreme items, or the distance of these items from the average, may also be computed as a supplement to the latter. Furthermore, the distance of the extreme items from each other or from the average is sometimes expressed as a percent of the average or of the highest or lowest item.[3]

[3] We must distinguish the method of characterizing a series given in its totality by the statement of the average and the extreme

The extremes of a series possess significance in that they give the " range " within which all observations of the series fall.[4] But knowledge of the " range " gives us no information of the distribution of the items within its limits. This distribution may vary widely even though maxima and minima are the same. On the other hand, series with different extremes may have practically the same conformation or dispersion. It is to be noted that the sizes of maxima and minima depend largely upon the number of observations. The greater the number of observations, the greater is the possibility of obtaining greater deviations from the average. Thus, the limits within which the heights of the inhabitants of a village fluctuate are not the standard limits for the whole country. Among the

values, from those cases where only the average and the maximum and the minimum (or one of these extreme values) are ascertained in the observation. Such cases are not rare. Thus, in the Austrian Labor Bureau's investigation of the condition of workmen in the Ostrau-Karwin coal district not only the detailed wages of the miners, but also, for purposes of comparison, wage data in small industrial establishments and the wages of agricultural and forest workmen were ascertained. That is, the amount of the highest, the lowest, and the normal, or average, cash wages of the workmen in question were asked for and classified by industry or nature of land. (Cf. Arbeiterverhältnisse im Ostrau-Karwiner Steinkohlenreviere, Pt. I, p. xvii.) In the same investigation the usual retail prices of certain commodities were also ascertained and, in case of great price fluctuations during the year, their lowest and highest prices.

[4] The extreme cases are sometimes defined more closely on account of their significance. If a phenomenon which changes in the course of time is characterized by the statement of its average and its maximum and minimum, then frequently the date is also given on which maximum and minimum have occurred (this is frequently done in the Tabellen zur Währungsstatistik, published by the Austrian Treasury Department). Often only one of the extreme cases is of significance. Thus, if factories, sick reliefs, etc., are arranged and presented according to size, the largest sick relief or the largest establishment, etc., are sometimes characterized by the statement of particular criteria (name, location, etc.).

larger population of the whole country deviations from the mean are probable which are entirely outside of the range for a single village. In spite of this, the distribution of the items about the mean in both cases may be essentially the same.[5]

Certain information about the dispersion in statistical series of quantitative individual observations may also be obtained by computing several averages instead of one. From the relative position of the various averages important conclusions about the conformation of the series may be drawn. If arithmetic mean, mode, and median coincide, or differ but slightly, then it is certain that the series is symmetrical. If they differ materially, it is of special importance to know whether the mode lies above or below the arithmetic mean or the median, and how much it differs from them. Concerning the relation between the median and the arithmetic mean, the fact that the arithmetic mean lies below the median indicates that the deviations below the median are greater than those above; *vice versa,* if the arithmetic mean lies above the median, the deviations above the median are evidently larger.

The median—found by bisecting the series—may be supplemented by stating those values which result from dividing the series into more than two groups of equal frequency. Thus, the quartiles originate from a division of the series into four groups of equal size. Stating the first and third quartiles (the second quartile being the median) of a series enables us to see within what limits half the items are located.[6] Furthermore, the deciles (or some

[5] Fechner, especially, has taken up the work on the connection between the number of observations and the sizes of the extreme deviations from the mean, and has developed laws of this relation for series which correspond to the normal or to the unsymmetrical Gaussian law. (Kollektivmasslehre, Chap. XX, "The Laws of the Extremes," pp. 321-338.)

[6] Thus, in the American special report, Employees and Wages (1903), in Chap. II, "Analysis of Occupational Comparison" (pp.

of them), found by dividing the series into ten equal groups, may be given. If, in a series, the mode of which is stated, other "secondary" points of concentration can be found, the statement of these is of special significance.[7]

A third way to give more information of a series which cannot be presented in detail than is possible by merely stating one mean, is to present certain characteristic classes —be they connected with the mean or independent of it— as well as the mean.

If, besides the median, other values originating from the division of the series into more than two equal groups be taken, such as the quartiles or the deciles, then they describe the classes adjoining the middle of the series, for the quartiles indicate between what limits below and above the median one-quarter of the items are located, the deciles indicate between what limits one, two, three, or four-tenths of the items are located above or below the median. Likewise, if we desire to supplement an arithmetic mean or a mode, we may compute and state within what limits are located one-quarter or one, two, three, or four-tenths, or any other fraction of the items.[8] The

xxix-xcix), the comparison of the wages of the workmen of various occupations is made for the years 1890 and 1900 by comparing the median and the two quartiles of each of the wage series referring to the two years mentioned. (Cf. ibid. p. xxviii.)

[7] Fechner (Kollektivmasslehre, Chap. XIX, "The Laws of Asymmetry," pp. 294-306) has developed a number of laws for the relation existing between the various means in series which correspond to the unsymmetrical Gaussian law. He has determined theoretically the intervals between these values, their relative positions, and the peculiarities of the deviations from the arithmetic mean and from the mode. Thus, he has determined a law of positions, according to which, if the unsymmetrical Gaussian law holds, the arithmetic mean and the median always lie on the same side of the mode in such a manner that the median falls between the arithmetic mean and the mode. (Cf. also Fechner, "Ausgangswert, etc.," p. 11.)

[8] Thus, March in "Quelques examples de distribution des salaires" (Journal de la Société de Statistique de Paris, 1898, p. 201) has

wider the limits between which a given fraction of items is located, the greater evidently is the dispersion. On the other hand, we may ascertain how many items are within definite limits above and below the average—for instance, how many items do not deviate more than 10% from the average. The greater the number of these items relatively, the closer the series will be condensed about the mean and the smaller its dispersion.

Aside from classes joining the average, classes independent of the average naturally can also be presented to supplement it. It may be especially significant to state the extreme classes—for instance, the extreme wage classes of a definite width together with the average wage. An even more concise characterization is obtained by merely stating the averages of the extreme classes. This is often better than the statement of the extreme cases which may be far removed from all the other items.[9] But then it must be mentioned, if possible, upon how many items the averages are based, since this cannot be deduced from the mere statement of the averages.[10] The most expedient way of supplementing the average can evidently be decided only from the peculiarities of the given case. The purpose

computed the intervals from the mode within which 30%, 50%, etc., of the wages which he is discussing are located.

[9] Therefore, the computation of the average of the maximum and the minimum, which we sometimes meet with, has only little value.

[10] Cf. Bowley, Elements of Statistics, 2nd ed., p. 93 ff. Bowley has also proposed to use the formula $\frac{Q_2 - Q_1}{Q_2 + Q_1}$ (in which Q_1 and Q_2 denote the two quartiles) for measuring the dispersion of series of quantitative individual observations, especially for purposes of comparison. This fraction increases with the distance between the two quartiles and clearly expresses changes in the series. Its value always lies between 0 and 1. By means of this formula Bowley has computed the dispersion of the wages of English mechanics for 1862 and 1890 and found the number 0 093 for the first year and 0.062 for the second year. Thus the dispersion of the wages had decreased. (Elements, p. 136.)

of the investigation in hand is the determining factor in the selection of any particular method.

Instead of gathering various details from the series to be characterized, in order to supplement the average, we may obtain a single numerical expression for the dispersion of the series by computing an *average of the deviations of the items from the average of the series* (*fluctuation number*). Recognizing the desirability of such a measure the Statistical Congress at The Hague in 1869 passed the following resolution: "Le congrès est d'avis qu'il est à désirer qu'on calcule non seulement les moyennes, mais aussi le nombre d'oscillations, afin de connaître la déviation moyenne des nombres d'une série de la moyenne de cette série même." However, according to G. von Mayr, it is not sufficient to compute the average deviation, but the latter must be expressed as a percentage of the average of the series.[11] But it does not appear to be always necessary to find this percentage. The absolute size of the average deviation may be significant and sufficient to supplement the average. Thus, the mean height of a given population may be given and, as a supplement to this average, it may be stated how many centimeters the individuals belonging to this population deviate *on the average* above or below this mean. According to G. von Mayr we ought to state the percentage that the average deviation bears to the average height.

The question, whether the average deviation should be stated as an absolute number or as a percentage of the average of the series, has been fully discussed by the mathematical statisticians. They are of different opinions. Lexis is decidedly opposed to the presentation of the probable and average deviations computed for typical series of measurements as a percent of the average of the series. In his opinion there is generally no reason for the assumption that the average (and also the probable) deviation depends

[11] Theoretische Statistik, p. 100.

in any way on the absolute size of the base value (the average of the series). "Suppose that the heights and the chest measures of a number of ten-year-old boys, and of a number of fully grown men have been taken. Presumably, the former series will result in a greater average deviation from the mean than the latter, and it is this difference which corresponds to the physical difference in the stability of the two anthropometric values. If we express the two deviations as percentages of the respective base values, then the divergence of these measures of dispersion becomes considerably greater than that of the absolute deviations; but the former cannot be compared with each other, while the latter are in an inverse proportion to the precision, and thus may be considered to be a direct and analogous expression of the dispersion."[12]

Fechner takes an essentially different stand. He distinguishes between repeated measurements of the same object affected with accidental errors of observation (physical and astronomic measurements) and measurements of various similar objects in a collective group. If series of the first kind are given, then the size of the deviation of the items from the mean is independent of the size of the mean (i. e., the object measured) and the absolute size of the average deviation must be stated. However, in series of the second kind it is Fechner's opinion that the deviations usually depend on the average size of the collective object in question,[13] and from this he draws the conclusion

[12] Abhandlungen, etc., VIII, "On the Theory of the Stability of Statistical Series," p. 173 f.

[13] "Generally speaking, errors of observation are essentially independent of the size of the object measured, at least with measurements of space and on the assumption that the measuring instruments are not changed. The errors of observation are larger when we measure a mile than when we measure a foot, it is true, but only because more and more complicated operations are necessary to measure the former, but the errors of observation when measuring a high thermometric or barometric registration are generally not

that with collective objects the average deviation must be expressed as a percent of the mean of the series, in order to eliminate the influence of its size.

Concerning the question, from what series may we compute measures of fluctuations, von Mayr (loc. cit.) says: "The measure of fluctuation can be computed arithmetically for any numerical series. It has a statistical value, however, only in certain series, particularly those in which the items have the character of similarity, especially in the sense that they present the process of social phenomena developing in equal time periods." Evidently von Mayr has time series primarily in mind. But even series

larger than when measuring a low one. The variations of collective objects, however, are essentially dependent on their sizes, if this be understood along the lines of the following illustrations: A flea is a small creature, and therefore the deviations of the individual fleas from the average flea are small on the average, being only fractions of the mean size, and the total difference between the largest and the smallest flea is but slight. The mouse is much larger than the flea, the horse again much larger than the mouse, a tree much larger than an herb, etc., and everywhere a corresponding observation is made, i. e., that the deviations of the individual mice from the average mouse are greater on the average than those of the individual fleas from the average flea, etc. This dependence of the average sizes of the deviations on the average sizes of the objects can also be explained by the fact that the interior and exterior modifying causes find more points of attack on large than on small objects. The quality of the object is also of significance because of the greater or smaller facility with which it yields to the modifying influence; and the accessibility to exterior modifying influences may differ with circumstances. Therefore, an exact proportionate relation of the average size of the deviations to the average size of the objects cannot be expected *a priori*. But in any case the sizes of the objects are principal factors influencing the sizes of their deviations." (Kollektivmasslehre, p. 77 f.) Also cf. Fechner, "Ausgangswert, etc., pp. 14 and 16. Georg Duncker differs from Fechner in regard to the dependence of the size of the deviations on the size of the objects measured. He denies the presence of such a dependence on the basis of his biological investigations. (Cf. Die Methode der Variationsstatistik, p. 40.)

of quantitative individual observations are not excluded from the measurement of dispersion by means of an average deviation. It is for just such series that mathematical statisticians have evolved their methods of measuring dispersion by the computation of various kinds of average deviations.[14]

The measure of fluctuation usually taken is the arithmetic mean of the deviations from the arithmetic mean of the series. However, other means may also be computed from the deviations of the items from the arithmetic mean of the series. Furthermore, the deviations of the items are not necessarily measured from the arithmetic mean of the series; on the contrary, some other mean of the series may be made the starting point. The choice of the mean of the series, from which the deviations are measured, and the choice of the measure of fluctuation which is computed from the deviations, are theoretically interdependent. The arithmetic mean of a series of items is characterized by the fact that the sum of the squares of the deviations from it is a minimum; the median, on the other hand, is characterized by the fact that the sum of the simple deviations of the items from the median is a minimum, i. e.,

[14] v. Mayr (Theoretische Statistik, p. 100) remarks that a knowledge of the deviations possesses the least objective interest for the "most pronounced form of typical series." By the "most pronounced form of typical series" v. Mayr (ibid. p. 90) means those series in which the arranged items lie symmetrically above and below the mean; the items are considered, so to speak, to be inaccurate reproductions of a constant base value, which in the actually observed phenomena is expressed with merely accidental deviations. However, it is not quite clear why the knowledge of the deviations of the items of such pronounced typical series should offer the smallest amount of objective interest, since even with symmetrical distribution of the items about the mean and with a central mode of the curve the sizes of the deviations may vary extremely. It is in the case of a symmetrical distribution of the items that the indices of fluctuation, as is explained later, are of the greatest scientific value.

smaller than the sum of the deviations of the items from any other value. Therefore, it would really be theoretically correct to base the average deviation (the arithmetic mean of the simple deviations) on the median of the series; if the arithmetic mean of a series is to be supplemented, then the mean square of the deviations ought to be used, which is found by dividing the sum of the squares of the deviations of the items by their number and then finding the square root of the quotient.[15] As a matter of fact, the mean square, as defined above, is generally, although not exclusively, used in the theory of error as a measure of the dispersion about the arithmetic mean of the series; it is called the *standard deviation.* Laplace has used the mean error of the simple deviations, although he measured these deviations in the usual way from the arithmetic mean of the series.[16] In elementary mathematical statistics the mean square is not known at all, and certainly never will be widely used for averaging the deviations from the arithmetic mean. Likewise the median of the deviations, i. e., that deviation which lies in the center of the series of deviations arranged according to size, the "probable" deviation which plays an important rôle in mathematical statistics, is rarely used in elementary mathematical statistics.

In order to be able to judge the methodological value of measures of dispersion we must keep in mind that while they offer, as averages of the deviations of the items from the mean of the series, a comprehensive numerical expression for these deviations, they *are* averages, and hence can never give a complete picture of these deviations and their sizes. This is a general truth whatever be the average of the series from which the deviations are measured, and whatever be the average which is computed from the deviations themselves.

[15] Cf. above, Pt. II, note 4.

[16] Cf. Fechner, "Ausgangswert, etc.," p. 54.

It is a peculiar quality of every measure of dispersion that in its computation really two groups of deviations are united and characterized by means of a single average. For there are the deviations of the items above the average of the series and the deviations of the items below this average. The numbers of the items above and below the average usually differ if we start from the arithmetic mean or the mode; if we start from the median, then, according to the definition of this average, the numbers of the positive and negative deviations are equal; but the sizes of the deviations above and below the median may vary. Therefore, with good reason, we could consider the positive deviations and the negative deviations of the items from the average of the series to be two separate series and compute the mean deviations separately for each set. However, this is not usually done, but the deviations of all the items are combined or united, so to speak, into a series, and this series is characterized by a single mean, the average deviation. If the positive deviations are equal to the negative deviations, then there is no objection to their combination and presentation by means of a single average (the mean deviation). But if these deviations are not equal, then, by combining them and computing a common mean, a value is obtained which correctly characterizes neither the deviations above nor those below the average of the series. Therefore the question is, when are the deviations of the items above the average of a statistical series equal to the deviations of the items below the average? Such a coincidence occurs when the whole statistical series is distributed symmetrically about the average from which the deviations of the items are measured.[17] Therefore, only in this case is the computation of a numerical measure of dispersion, combining all the deviations from the mean into a single average, entirely free from objection. This

[17] It is not necessary that the series contain a central mode as the theory of error presupposes, but it is usually the case.

measure of dispersion naturally varies according to the distribution of the items on both sides of the mean. If the items of the series are not distributed symmetrically about the mean of the series from which the deviations are measured, then the measure of dispersion computed for the series has only little value and may actually lead to error. Let us imagine a series which has a mode on one side of the arithmetic mean—but not very far from it—while on the other side of the arithmetic mean fewer, but very marked, deviations are present, so that the conformation of the series above and below the arithmetic mean is not at all the same. If, starting from the arithmetic mean, we compute a measure of dispersion for this series, we obtain a small number and, merely knowing this number, we are likely to assume that the whole series is condensed about the arithmetic mean with little dispersion.[17a]

Therefore, with series of irregular conformation it is better not to take an average of all the deviations as a measure of dispersion. But, under certain circumstances, it may be both allowable and expedient to compute separate measures of dispersion for the items above and below the averages of such series. This is advisable if the structure of the series differs on the two sides of the base value while each side, taken alone, shows a regular conformation. Such a dispersion is found quite frequently if the mode of a series

[17a] The author is not in accord with the general tendency in laying so much stress on symmetry as a prerequisite for the use of measures of dispersion. Too great insistence that statistical data must conform to mathematical rules is fatal; it destroys the usefulness of the science of statistics. Even though series are not symmetrical, measures of dispersion may be profitably employed to characterize them. Of course, in computing the arithmetic average deviation from the arithmetic average of a series all deviations are considered positive; in computing the standard deviation the squares of the deviations only contribute to the size of the resulting measure of dispersion.—TRANSLATOR.

of individual observations is made the base value. In this case we may try to express the deviations of the parts of the series on both sides of the modes by separate averages of deviations. If not even the conditions for this kind of expression be given, then we must try to characterize the dispersion of the series as well as possible in some other way: by stating the extremes, by computing various averages, by presenting certain classes, by stating how many items lie above and how many below the arithmetic mean or the mode, etc. Finally, it must be noted that there are series of items which are distributed about the mean with no regularity, but which have some other regular characteristic structure. The nature of such a series cannot be ascertained by investigating the dispersion of the series about its mean, but the peculiar regular conformation must be ascertained and expressed by other methods which are discussed in Appendix I.

It must be mentioned in this connection that the kind of dispersion of series of quantitative individual observations essentially depends on whether or not the series is homogeneous. Series of a heterogeneous make-up usually show an irregular conformation. They generally contain several points of concentration of equal or varying importance—which correspond to the modes of the combined constituents—and there is no symmetric distribution about any mean. If such series are decomposed into the more homogeneous constituents of which they consist, then constituent series result, each of which frequently shows only one mode, centrally located, which approximately coincides with the arithmetic mean and the median; the items being distributed with a certain symmetry and regularity about the means of each homogeneous constituent series. If the wages of all the workmen of a given district are combined in a statistical series, a very irregular conformation is usually the result. Since sex generally has a decisive influence on the amount of wages, the total

series, giving the wages of men and women, probably contains two points of concentration for the two sexes. Even if men and women are treated separately several points of concentration may occur, caused by combining workmen of different occupations. In addition, the wages of the skilled and unskilled laborers of the same occupation may stand forth as separate groups.[17b] But if all the elements of differentiation influencing larger groups of workmen are taken account of, and if essentially homogeneous constituent series are formed by suitably decomposing the entire series into constituent series which contain only individual differences, then the occurrence of several points of concentration is not probable, and we can count on a symmetric distribution.

Besides attempting to secure greater homogeneity of kind we may also strive for greater homogeneity of space and time and, in this way, obtain an improvement of the conformation of the series. Items originating in different countries or districts frequently give rise to varying averages and, therefore, in case of their combination, to several points of concentration which may be removed by suitably decomposing the series. Measurements of the stature of individuals of various nationalities result in an irregular series, while the stature of the inhabitants of the same country are distributed regularly about their average.

However, there may also occur homogeneous series of unsymmetrical conformation. In biology and anthropology such series generally indicate an evolution, to be considered an improvement or a degeneration according to the nature of the case. The fact that evolution is the cause of an

[17b] Likewise, Prof. H. L. Moore has shown that union and non-union workmen belong in separate categories as regards wages (see Laws of Wages, p. 189). Age of employees and size of establishment also appear to give significant wage categories (Laws of Wages, p. 143).—TRANSLATOR.

unsymmetrical conformation is explained by Lexis[18] in the following way: A certain percentage, say, half of the totality, is still distributed regularly about the original type, while the rest shows another distribution on account of external influences or other causes. If there is, for example, degeneration, caused by injurious influences, we may assume that the subnormal groups are affected by these influences to an extent measured by their distance below the normal type and that the supra-normal groups are influenced comparatively less. By combining the degenerate half with the stable half an unsymmetrical distribution of the totality results, in which groups below the normal are more extended than the groups above the normal.

B. MEASUREMENT OF THE DISPERSION OF SERIES OF QUANTITATIVE INDIVIDUAL OBSERVATIONS FROM THE STANDPOINT OF THE THEORY OF ERROR

When judging the dispersion of series of quantitative individual observations from the standpoint of the theory of error the point at issue is to ascertain whether the distribution of the items corresponds to the normal (symmetrical) law of error as defined by the Gaussian law, or to any generalized (unsymmetrical) law of error.

The normal law of error was established and formulated originally on the basis of repeated observations of the same object, especially such as repeated measurements of the same object in the astronomical, physical, or geodetic fields. We know empirically that repeated measurements of an object, the size of which is to be ascertained, do not completely coincide. The various measurements are affected with varying "accidental" errors of observation. However, the series formed by the individual measurements shows a regular characteristic conformation. The series is

[18] Cf. "Anthropologie und Anthropometrie" in the Handw. d. Staatsw., 2nd ed., Vol. I, p. 397, and Abhandlungen, etc., p. 124.

distributed symmetrically about the arithmetic mean, the most probable value of the object measured, the items being crowded most densely around the arithmetic mean and becoming rarer the farther distant they are from it. The frequency of the various measurements is a function of their distance from the arithmetic mean or, in other words, the frequency of the varying deviations from the arithmetic mean is a function of the size of these deviations. Gauss was the first to investigate the probability of the varying deviations, and has formulated the law of distribution (called after him the Gaussian probability integral) for the dispersion of the items about the arithmetic mean.[18a]

The sizes of the deviations in a series of repeated measurements of the same object depend on the precision of the individual measurements. The greater this precision, the smaller are the accidental errors and the more densely are the items crowded together. Consequently, the graphic presentation of the series results in a curve which, with great precision of the items, extends only over a small part of the axis of the abscissas and declines abruptly on both sides of the arithmetic mean, while the curve which is based upon less precise measurements is more extended, and the items show greater deviations on both sides of the average.

The dispersion of a series of repeated measurements of the same object, obeying the Gaussian law, can be characterized by a single expression obtained by computing an average of the deviations from the arithmetic mean. As such measures of dispersion the error of mean square (standard deviation), the average error, and the probable error as well as the modulus may be used.[19] Certain

[18a] See the equation of the Gaussian curve on p. 166, footnote 37a.

[19] The modulus is to be computed as the square root of twice the quotient of the sum of the deviations from the arithmetic mean divided by the number of items. Edgeworth has proposed the term

mathematical relations exist between these measures of dispersion, so that one can be computed from the other.[20]

A picture of the dispersion of the items of a symmetrical series, obeying the normal law of error, about the arithmetic mean of the series is given in the following table taken from a paper by W. Townsend Porter,[21] a table which shows plainly the connection between the number of items and their deviations from the mean. If M denotes the arithmetic mean, d the probable error, which may vary in size in given cases according to the precision of the individual measurements, then 1,000 items will be distributed about the mean in the following way:

	+nd	3
	+4d	18
	+3d	67
	+2d	162
	+d	250[22]
M		
	—d	250[22]
	—2d	162
	—3d	67
	—4d	18
	—nd	3

"fluctuation" for the square of the modulus. The reciprocal value of the modulus is the "precision," which may also be used as a measure of dispersion. (See footnote 37a, p. 166.)

[20] Cf. especially Bowley, Elements of Statistics, 2nd ed., pp. 281-292; Fechner, Kollektivmasslehre, pp. 18-22, and Duncker, Die Methode der Variationsstatistik, p. 36 f.

[21] "On the Application to Individual School Children of the Means Derived from Anthropological Measurements by the Generalizing Method." (Bull. de l'Inst. intern. de Stat., Tome VIII, 1895, p. 279 f.)

[22] That one-quarter of the cases lie within the probable error below the average and one-quarter above the average is a direct consequence of the definition of the probable error.

The greater the distance from the mean the smaller becomes the number of items in a definite proportion as defined by the Gaussian probability integral.

A similar, but more detailed table for the distribution of 1,000 measurements affected with accidental errors is given by Colajanni in his *Statistica Teorica* (p. 192), as follows:

SIZE OF THE ERROR (EXPRESSED BY THE PROBABLE ERROR AS UNIT)	NUMBER OF DEVIATIONS		
	POSITIVE	NEGATIVE	TOTAL
From 0 to 0.5	132	132	264.1
" 0 5 " 1	118	118	235.9
" 1 " 1.5	94 2	94 2	188.3
" 1.5 " 2	67 2	67 2	134 3
" 2 " 2 5	42 8	42 8	85.6
" 2.5 " 3	24 4	24.4	48 7
" 3 " 3.5	12 4	12 4	24 8
" 3.5 " 4	5 6	5 6	11 2
Over 4	3.5	3 5	7.1
	500	500	1000

The attempt has frequently been made to explain the characteristic regular conformation of the series resulting from repeated measurements of the same object. It is generally supposed that this conformation can be traced back to the presence of a great number of independent, equally positive and negative sources of error (contributory causes). Every one of these sources of error, taken by itself, can cause only a very small positive or negative error. The single measurements, however, are always influenced by several sources of error simultaneously, and that in various combinations, causing greater and smaller deviations. In order to cause a greater deviation, either in the positive or the negative direction, several sources of error acting in the same direction must operate simultaneously or, if positive and negative sources of error occur together, one side must outweigh the other considerably,

since positive and negative sources of error occurring in equal numbers counterbalance each other. Therefore, greater deviations from the average (corresponding to the true size of the object measured) can only result from certain combinations of the sources of error, which combinations possess probabilities decreasing with the sizes of the deviations. From this results the fact, which originally was ascertained empirically, that in a series of repeated measurements of the same object the items become rarer the farther they diverge from the mean.[23]

It is not always easy to define the nature of the sources of error in a given case of concrete measurements. With physical measurements the inadequacy of the human organs of sense, especially of the eye, the uncertainty of the hand, the imperfections of the instruments used, etc., may be considered to be the sources of error.

Series which obey the normal law of errors originate not only from repeated measurements of the same object, but sometimes from single (statistical) observations of distinct, but similar items. Therefore, mathematical statisticians usually examine statistical series in order to ascertain whether or not they agree with the normal law of error.[24] If there is sufficient evidence for this, then the series is "typical," i. e., a definite normal value is expressed in the series with merely accidental deviations. The mathematical theorems formulated originally for repeated measurements of the same object can then be applied with good reason to the statistical series in question.[25] Especially the

[23] Only an immaterial modification of the above hypothesis is found if we follow Pearson and do not proceed from the assumption that equal numbers of positive and negative sources of error (contributory causes) are present, but from the assumption that every source of error acts positively or negatively with the same probability.

[24] Cf. for the methods which may be applied to this investigation, Czuber, Wahrscheinlichkeitsrechnung, pp. 335-341.

[25] With Lexis we may well speak in this case of a "physical"

dispersion of the series can then be measured and expressed according to the theory of errors of observation by a single measure of dispersion, such as the error of mean square, the probable error, the average error, the "modulus," or the "precision." The errors are measured from the arithmetic mean of the series which may be called the "typical" mean. But since this mean is located centrally and since the items are densest about it, the arithmetic mean, mode, and median coincide, or differ slightly on account of the small number of observations. If the coincidence of the series with the normal law of error is established and if the average and the average dispersion of the series are known, then the series is completely characterized for the mathematical statistician. He knows the law of distribution to which the series corresponds, and the constants applying to the individual series.

It may also occur that, although the entire series does not coincide with the law of error, a portion of it corresponds to this law and shows a typical conformation. In such a case the typical mean is not expressed by the average of the entire series, but by that value about which that part of the series which agrees with the law of errors is distributed. For the rest, analogous consequences are found as in the case of wholly typical series.

The analogy between "typical" statistical series and series of repeated measurements of the same object also leads to a plausible explanation of the former. For it may be conceived that typical statistical series originate from the action of a great number of independent causes working, either in positive or negative direction and in various combinations, upon the individuals or items observed and expressed in the series. Thus, the conformations of the series of measurements of human heights which, as we know, frequently correspond to the symmetrical law of

method, since a method used primarily in physical observations is applied here in statistical material.

error, may be traced back to the fact that numerous influences (such as different nutrition, various modes of living in adolescence, atavism, etc.), partly promoting growth, partly retarding it, work upon the various individuals in varying combinations, thus causing a symmetrical distribution about the average height, corresponding to the law of error.[26]

However, " typical " statistical series corresponding to the law of error occur but rarely. Quetelet has proved merely that they exist in the field of anthropometry. He found a symmetric distribution about the average, corresponding to the law of chance, especially in series of measurements of height and girth of chest. In his investigations Quetelet did not use the Gaussian law, but the binomial formula, in which the probabilities of the various deviations from the average correspond to the coefficients of the binomial expansion.[27] Quetelet's expectation that series corresponding to the symmetrical law of chance would frequently be met with outside of the field of anthropometric statistics has not materialized. We know to-day that the normal law of error cannot be taken at all as the general law of distribution of statistical phenomena. Statistical series or part-series corresponding to this law of distribution are found only in very rare cases outside of the anthropometric field. The best known of these cases, discovered by Lexis, is the symmetrical distribution of the items of the mortality table about the " normal " length of life, the mode of the series of lifetimes given in the mortality table. But this symmetrical distribution extends only to certain age classes belonging to the " normal group " which includes the normal length of life. This is a comparatively small part of the total series of lifetimes

[26] Cf. Lexis, "Anthropologie und Anthropometrie" in Handw. der Staatsw., 2nd ed., Vol. I, p. 389 f.

[27] Cf. about Quetelet's binomial table in the work quoted in note, p. 390 ff.

which in its entirety does not correspond at all to the law of error, a fact which is expressed by the strong divergence of the arithmetic mean from the mode of this series.[28-29]

But even in the field of anthropometry the normal law of error cannot be taken as the universal law of distribution. Anthropometric series frequently show unsymmetrical distributions (for instance, measurements of weights), sometimes contain several points of concentration, which are caused by lack of homogeneity, and sometimes show no perceptible point of concentration.

Statistics, therefore, usually has to do with series which neither in whole nor in part (as in the case of the normal length of life) admit the application of the methods of the theory of errors of observation in their original

[28] Even the symmetry of the "normal group" of lifetimes has already been attacked by Pearson on the basis of English material, in "Contributions to the Mathematical Theory of Evolution," II: "Skew Variations in Homogeneous Material." Philosophical Transactions of the Royal Society of London, Vol. CLXXXV, 1, 1895, A., p. 407.) New investigations of the distribution of the items about the mode (normal value) have been published by E. Blaschke in Vorlesungen über math. Statistik (p. 154 ff.). These investigations refer to various mortality tables, especially mortality tables of insurance companies, invalidity tables, and to several series giving the number of sick days according to age, and the distribution of those marrying according to age. Blaschke (in substance agreeing with Lexis) found that the distribution of the items above the "normal age" (towards the older age classes) approximately corresponds, in most cases, to the law of distribution of errors, but that below the normal age the coincidence covers only a narrow range.

[29] A remarkable coincidence with the theory of error is also shown by the fluctuations of the net proceeds of various cereal crops in Germany. The fluctuations of the net proceeds were computed for some successive cereal crops by means of the probability calculus ("Die Schwankungen, etc.," by Dr. Alfred Mitscherlich, Supplement No. VIII of the Zeitschrift für d. ges. Staatsw., Tübingen, 1903), and it was shown to what important practical results the application of the theory of error may lead in forming precepts for agricultural management.

form and, consequently, cannot be characterized sufficiently by a single measure of dispersion.

With this fact in mind mathematical statisticians have tried to express and explain mathematically the conformation of non-symmetrical series by modifications and generalizations of the Gaussian law. They have made it their purpose to establish laws of distribution for non-symmetrical series which shall enable them to combine the various statistical series into theoretically defined groups of a higher order on the basis of uniform principles.

The reason for accounting for non-symmetrical statistical series by the use of some appropriate extension of the law of error is much more obvious when we consider how many unsymmetrical statistical series exhibit only little asymmetry, while their structure otherwise closely approaches that of really " typical " series. The transition from the symmetrical series which correspond to the normal law of error, to the undoubtedly unsymmetrical series is indefinable, since complete symmetry never occurs on account of the usually relatively small number of observations. Therefore, in a given case opinions may differ as to whether the asymmetry of a definite series is to be considered unessential and caused merely by the insufficient number of observations, or as essential, so that an unsymmetrical law of distribution is expressed in the dispersion of the items. From the unsymmetrical but regular statistical series innumerable forms of transition lead to conformations which appear to possess no regularity and which seem to exclude a theoretical explanation. However, mathematical statisticians have frequently tried to bring even such series under a generalized law of error and have endeavored to prove that this universal law is expressed even in these series, although with considerable modification.

The main representative of this idea is Edgeworth. In his opinion preference must always be given in the mathematical presentation of statistical series to such formulæ

as have a certain relation to the normal law of error and merely show deviations from it. The series described by these formulæ are supposed to be caused by certain modifications of the conditions under which, *otherwise*, the normal law of error would originate.[30] Edgeworth does not consider it to be the most important point that the statistical material correspond to the highest possible degree with the formula chosen to present it, but he insists that the chosen law of distribution be based on a hypothesis which is plausible *a priori*, and that this law offer a plausible explanation for the distributions under examination. All this holds in Edgeworth's opinion for the law of error and, consequently, the examination of statistical series must start from it. Therefore, the representation of statistical series by means of the law of error and appropriate extensions of it, even if the theory does not completely agree with the material of observation, is more valuable than the representation by means of an empirical formula (analytic function) which, even though it fits the material perfectly, does not furnish any explanation of the conformation of the series at hand.[31]

In the first place we may try to decompose unsymmetrical series into constituent series which separately correspond to the normal law of error (method of separation and unsymmetrical Gaussian law) or to reduce them hypothetically to an original conformation which corresponds to this law (method of translation). If we succeed in doing this, the further mathematical treatment is accomplished by the methods of the normal theory of error. To this first group of methods are opposed those methods in which we do not stop with the normal theory of error but explain dis-

[30] On the "Representation of Statistics by Mathematical Formulæ," Journ. of the Roy. Stat. Soc., 1898, p. 674.

[31] The presentation of statistical series by means of empirical formulæ is discussed only in Appendix I, since this manner of presentation does not proceed from the average.

tributions by means of skew curves of error as well as by means of other generalizations of the law of error or the binomial law.

The method of separation. The idea of considering multimodal curves as complex curves caused by addition or subtraction of two or more curves, is obvious even to non-mathematical statisticians. In this way the two vertices in the frequency curve of the heights of the recruits of certain French departments have been traced by J. Bertillon to the mixture of two races of different type as to height. This process is more difficult with unimodal curves which are supposed to be made up of two superimposed modes. Pearson,[32] especially, has taken up the problem of decomposing such curves in two constituent curves of normal dispersion. He did not limit himself to unsymmetrical curves, but declared that even the decomposition of symmetrical curves was sometimes necessary. Unsymmetrical as well as symmetrical curves may consist of more homogeneous constituents of normal dispersion, the separation of which is of scientific value. Furthermore, the decomposition of an unsymmetrical curve which does not consist of heterogeneous material, may be expedient, since by this decomposition we obtain a measure of the irregularity of the curve and, if we are working in the field of biology or anthropology, a measure of the evolution of the given character, which may express an improvement or a degeneration of the species.

Pearson has[33] also promised an investigation of the problem of decomposing complex curves into skew constituent curves and,[34] by way of experiment, has decomposed

[32] "Contributions to the Mathematical Theory of Evolution" in Philosophical Transactions of the Roy. Soc. of London, Vol. CLXXXV, 1894, A, pp. 71-110. See also Duncker, Die Methode der Variationsstatistik, p. 16 f.

[33] Philosophical Transactions of the Roy. Soc. of London, Vol. CLXXXVI, 1895, Pt. I, A, p. 406. [34] Ibid. pp. 406-410.

the series of lifetimes given in the English mortality table into 5 constituent curves, whose modes are at the ages of 71.5 years, 41.5 years, 22.5 years, 3 years, and in the beginning of the first year, corresponding respectively to the mortalities of old age, of middle age, of adolescence, of childhood, and of infancy. Of these five constituent curves, those of the mortalities of old age, childhood, and of infancy are unsymmetrical, those of middle age and of adolescence are almost symmetrical. It is remarkable that Pearson has found an unsymmetrical distribution for the mortality of old age, at least on the basis of the English material, while Lexis has ascertained a symmetrical distribution about the maximum, the normal length of life, on the basis of the material of various other countries.

The decomposition of series according to the method of separation may enable us to draw conclusions of great significance, as the following example shows. In his book *Die natürliche Auslese beim Menschen,* Ammon has based his assertion of an evolution of the cephalic index of the South German on the comparison of exhumed Germanic skulls with modern skulls. In opposition to this, Pearson states in the paper quoted above that he has succeeded in decomposing the unsymmetric curve of the old Germanic skulls into two constituent curves, one of which agrees with the curve found for modern skulls. From this, Pearson concludes that the older skulls come from a mixed population, and that an evolution of the cephalic index cannot be proved for that part of the population which is still living.

Pearson's method has been criticised thoroughly by Edgeworth,[35] who acknowledges that Pearson's method of separation—in spite of its mathematical complexity—is very valuable, especially because it is not based on a merely theoretical hypothesis but on actual facts. It is, for instance, known that the statistical series representing the

[35] Journ. of the Roy. Stat. Soc., Vol. LXII (1899), p. 125; see also Vol. LXV (1902), p. 327.

heights of the Italian population can be divided into constituent series for the various provinces, which show averages of different sizes and represent different types of height.

The unsymmetrical Gaussian law. An unsymmetrical series corresponds to this law if the two parts of the series located on either side of the mode correspond *separately* to the normal Gaussian law and if the non-symmetry of the series is caused merely by the fact that the two parts of the series have different mean errors (or probable errors, moduli, etc., according to the measure of dispersion used). According to the unsymmetrical Gaussian law, unsymmetrical series are considered to be complex series consisting of two half-normal curves of different precision joined at their modes. Obviously, only those series which possess but one mode may be so considered. If a series is found for which the above assumption holds, then it can be completely defined mathematically by the statement of the mode and the mean error (or any other measure of dispersion) of each of the two constituent series.

Professor Edgeworth, who has thoroughly explained and discussed the method of the non-symmetrical Gaussian law under the heading "Method of Composition,"[36] admits that, as a matter of fact, this method can sometimes be used with success in the mathematical treatment of certain series. But he also states that very unsymmetrical series cannot easily be treated by means of this method because a curve of regular conformation cannot originate from the halves of two normal curves of considerably varying dispersion. Edgeworth also objects to the "method of composition" because it contains no plausible reason for the assumed kind of conformation of the series. Why the constituent series on either side of the mode should contain different mean errors, is not explained at all by the assump-

[36] Journ. of the Roy. Stat. Soc., Vol. LXII (1899), pp. 373-385 and p. 543.

tions used in this method, and it is not clear why such an artificial and manufactured conformation of the series should originate.

The method of the unsymmetrical Gaussian curve has received the most thorough treatment by G. Th. Fechner. In his paper published in 1878, "Über den Ausgangswert der kleinsten Abweichungssumme," Fechner originally presented the idea of considering unsymmetrical series of measurements as complex curves consisting of halves of normal curves of errors of different precisions, and he has developed the idea most exhaustively in his *Kollektivmasslehre* (1897). In this work Fechner has given a number of series which agree with the unsymmetrical Gaussian law and which, on the basis of his investigations, justify the assumption that this law must be considered to be the universal law of distribution of "collective objects"—a term which Fechner uses to designate all accidentally varying objects, especially anthropological, biological, and meteorological measurements. It is, however, limited to series with relatively weak fluctuations about the mode, such as appear in most collective objects.

In the case of great asymmetry and great relative deviations Fechner has recommended the "logarithmic" treatment of the given series.[37] This method consists in finding the logarithms of the items, determining the mode in the series of these logarithms, and examining the deviations of the individual logarithms from their mode. By this method Fechner has arrived at a logarithmic generalization of the unsymmetrical Gaussian law, inasmuch as he found that this law of distribution holds also for the series of logarithms which he examined.

According to Fechner, the logarithmic treatment is valuable principally because, by translating the logarithmic mode and the logarithmic deviations into the appertaining natural

[37] Cf. Kollektivmasslehre, pp. 77-83 and pp. 339-351, and "Ausgangswert, etc.," pp. 14-17.

numbers, we obtain the relative deviations and their base value, the relative mode. The latter differs from the arithmetic mode. The relative deviation indicates the *percent* of the base value that the item is above or below it, while the usual arithmetic deviation indicates the absolute difference between the item and the mean. According to Fechner, the relative deviations have a special significance in the field of collective objects—but not in physical and astronomical observations—since "the variation of an object bears a certain relation to the size of the object itself, so that the variation depends essentially, although not exclusively, upon the size of the object; according to this, the height of a blade of grass, taken absolutely, varies less than that of a fir tree, but we cannot assert that relatively it varies less."[38] Therefore, according to Fechner, the variation of a collective object can be judged better from relative deviations than from arithmetic deviations. Furthermore, the fact that arithmetic deviations are limited, inasmuch as an object cannot decrease more than its own size, argues for the logarithmic treatment and the computation of relative deviations, because this limitation is removed when referring to logarithmic and the ensuing relative deviations, since any object may decrease as well as increase in infinite ratio.[39]

In Fechner's opinion the normal (symmetric) Gaussian law based on arithmetic deviations is merely a special case of the unsymmetric Gaussian law also based on arithmetic deviations. It corresponds to the limiting case where the deviations on both sides of the mode are equal. Among the infinitely many degrees of varying asymmetry the case of the complete disappearance of asymmetry possesses only little probability. On the other hand, the logarithmic law of distribution, which is to be applied to collective objects

[38] "Ausgangswert, etc.," p. 14; see also Kollektivmasslehre, p. 78 f.; and above, p. 262 f.

[39] "Ausgangswert, etc.," p. 16, and Kollektivmasslehre, p. 77.

with strong relative fluctuations, agrees remarkably, according to Fechner, with the unsymmetrical Gaussian law based on arithmetic deviations if there are only weak relative fluctuations, and becomes identical with it as the fluctuations decrease, so that the logarithmic law may be taken as the most general law of distribution of collective objects.

Fechner has also tried to explain the origin of unsymmetrical series which correspond to the unsymmetrical Gaussian law.[40] He assumes the existence of an indefinite number of forces or special conditions which, independent of each other, exert a modifying influence upon the sizes of the specimens of a collective object. According to Fechner's hypothesis, every force, when active, causes so small a change that the second power of the latter may be neglected in comparison with finite values. There is the probability p for the presence of an effect, and the probability $q = 1 - p$ for the absence of an effect of the activity of each individual force. On the basis of this hypothesis an unsymmetrical distribution is generally obtained; a symmetrical distribution corresponding to the normal Gaussian law originates merely in the special case, when p equals q.

The method of translation. Edgeworth, the originator of this method, proceeds from the following observation:[41] If a series is given which corresponds to the normal symmetrical law of error and if, based on this series, a second series be computed the items of which represent assigned functions of the items of the first series, then the second, generated series, under certain conditions, has an unsymmetrical conformation which is mathematically predetermined. Let the original series be the distribution, corresponding to the normal law of error, of the heights of the males of any nation. Now, if a second series is formed from, say, the squares of those values in which the heights

[40] Kollektivmasslehre, pp. 306-320 and p. 357.

[41] Cf. Journ. of the Roy. Stat. Soc., 1898, p. 675, and 1899, p. 537.

of the first series result when taken from a definite point,—for instance, from the height of the shortest man,—then this new generated series shows an unsymmetrical conformation. According to Edgeworth, the essence of the method of translation is to treat unsymmetrical series of measurements as though every item were an assigned function of an item of a normal series which is to be ascertained. The problem, then, is: to find the generating normal curve, which may be determined by computing the average and the modulus. An additional problem is: to determine a point such that its distance from each point of the generating curve is in functional relation with a corresponding point of the generated curve. With the help of some examples taken from the field of meteorology (especially unsymmetrical series of barometric measurements) Edgeworth has shown how this method can be applied in practice.

The method of the skew curves of error. The aim of the methods mentioned thus far is to divide unsymmetrical series into symmetrical series or constituent series or to reduce unsymmetrical series to symmetrical series in order to apply a thorough mathematical investigation to the symmetrical series or constituent series thus obtained. Instead of proceeding in this way we may proceed with a direct mathematical treatment of the unsymmetrical series. The problem, then, is to find a skew curve of error with which the actual material of the series agrees sufficiently so that this curve or the corresponding algebraic function may be used in expressing the series in question. Of course stating an average of the deviations from the arithmetic mean of the series is obviously not sufficient to completely characterize a skew curve of error, but a numerical expression of the degree of asymmetry must also be computed.[42] How-

[42] Cf. Edgeworth, on the "Asymmetrical Probability-Curve," Philosophical Magazine, 1896; and Journ. of the Roy. Stat. Soc., 1900, p. 76; and Bowley, Elem. of Stat., 2nd ed., Appendix, pp. 329-334; and Journ. of the Roy. Stat. Soc., 1902, pp. 331-354.

ever, only series which do not deviate too much from the normal form can be reproduced by means of skew curves of error.[43]

The generalized method of translation. Another method belonging in this list is the generalized method of translation, developed by Professor Edgeworth.[44] Its most general application is to relate a given series to an unsymmetrical generating curve. By means of this method Edgeworth has succeeded in expressing mathematically even series of strongly unsymmetrical conformation, as well as unilateral curves the modes of which are in the beginning or at the end of the series. Such unilateral curves frequently occur in botanical statistics. But they also occur in population and economic statistics. Such series are, principally, the childhood mortality which begins with a maximum, and the series of taxable incomes, in which the lowest classes are of the greatest frequencies, at least in some countries. Edgeworth admits that such unilateral series, especially the distribution of incomes, apparently have nothing to do with the law of error.[45] But the income may be in functional relation to some other criterion which is governed by the law of accidental deviations, such as individual ability.[46] Therefore, Edgeworth thinks it is admissible to

[43] Especially Thiele, Bruns, Charlier, and Kapteyn have worked on the skew curves of error from a theoretical mathematical point of view.

[44] Cf. Journ. of the Roy. Stat. Soc., 1899, p. 537.

[45] The attempts made outside of the field of theory of error to characterize such series by means of analytical formulæ (which do not originate from the averages of the series in question) are discussed in Appendix I (A).

[46] According to the curve drawn by Galton, the various degrees of human faculty are distributed about the mean according to the Gaussian law. (Cf. especially Hereditary Genius and Inquiries into Human Faculty and Its Development.) Cf. also Ammon's discussions (Die Gesellschaftsordnung und ihre natürlichen Grundlagen) on the relations between the distribution of incomes and the curve of faculties.

treat the distribution of incomes and similar unilateral series by means of a generalized method of translation, and believes himself to have obtained in this way better results, i. e., a better agreement between theory and material of observation, than Pearson, who also has treated series of this kind by means of a special generalization of the law of error (discussed later).

The "generalized law of error" of Edgeworth. The law of error has recently been given its most general form by Edgeworth, whose "generalized law of error" or "exponential law of great numbers"[47] is adapted to explain the widest range of statistical series. The normal law of error as well as the method of translation are only special cases of this generalized law, the validity of which can therefore be proved in many fields of nature and social life when they are treated statistically.

Pearson's generalized probability curve. Finally, the generalized probability curve must be mentioned. It was formulated by Pearson and he has illustrated it by numerous examples taken from various fields.[48]

It corresponds to the general binomial $(\frac{p}{p+q}+\frac{q}{p+q})^n$ in a similar way as the normal curve (Gaussian curve of error) does with the definite binomial $(\frac{1}{2}+\frac{1}{2})^n$.[49] It may therefore, be symmetrical or unsymmetrical and of unlimited

[47] Cf. Journ. of the Roy. Stat. Soc., 1906, p. 497 ff.

[48] See "Contributions to the Mathematical Theory of Evolution," II: "Skew Variation in Homogeneous Material" (Philosophical Transactions of the Royal Society of London, Vol. CLXXXVI, Pt. I (1895), A., pp. 343-414) and divers articles in Biometrika, a journal for the statistical study of biological problems, edited, in consultation with Francis Galton, by F. R. Weldon, Karl Pearson, and C. R. Davenport. (Cambridge, since October, 1901.) Cf. also C. B. Davenport, Statistical Methods with Special Reference to Biological Variation, New York, 1904, and W. P. Elderton, Frequency-Curves and Correlation, London, 1906.

[49] Pearson, loc. cit., p. 345, and Duncker, Die Methode der Variationsstatistik, p. 14.

extension on both sides of the mean on the axis of the abscissas, or limited on one side or on both sides of the mean. Thus, the following five types are found:

TYPE 1. Axis of the abscissas limited on both sides, curve unsymmetrical.

TYPE 2. Axis of the abscissas limited on both sides, curve symmetrical.

TYPE 3. Axis of the abscissas limited on one side, curve unsymmetrical.

TYPE 4. Axis of the abscissas unlimited on both sides, curve unsymmetrical.

TYPE 5. Axis of the abscissas unlimited on both sides, curve symmetrical. Type 5 is the normal curve or Gaussian curve of error.[50]

An unsymmetrical conformation of a given curve can be explained, according to Pearson, in that not all the contributory causes bring about equally great positive or negative deviations with the same probability, which is, according to Pearson, to be assumed for the origin of a normal curve of error.[51] Furthermore, Pearson's probability curve is, in its most general form, based on the assumption that the contributory causes are not independent of each other.[52]

[50] Pearson, loc. cit., p. 360; Duncker, loc. cit., p. 15 f. (Cf. the classification of curves given by W. Palin Elderton in Frequency-Curves and Correlation.—TRANSLATOR.)

[51] If we do not proceed like Pearson, when explaining the symmetrical curve of error, from the assumption that the contributory causes may with equal probability cause equally great positive and negative deviations, but from the more usual assumption that a symmetrical curve of error must be based on the activity of two equally strong groups of contributory causes, one of them causing positive and the other negative deviations, then we may explain the origin of unsymmetrical curves of error by the assumption that the contributory causes acting positively and negatively in producing individual variation are not present in equal numbers. (Cf. Duncker, Die Methode der Variationsstatistik, p. 33.)

[52] Cf. in this connection the criticism by Edgeworth in Jour. of the Roy. Stat. Soc., 1899, p. 535 f.

Great stress is put by Pearson on the fact that the range of variation of most phenomena is limited, owing to their nature, a fact which is theoretically opposed to the application of the normal curve of error which is unlimited on both sides. Thus, every age distribution has a fixed inferior limit and, usually, also a superior limit; for instance, the age distribution of the women who give birth to children during the year is limited on one side by the age of puberty, on the other side by the climacteric age.[53]

The practical application of Pearson's theory consists, above all, in computing the function to which a concrete statistical series corresponds. In addition, the mean of the series, the standard deviation, the number of observations contained in the series, and a measure for the degree of agreement between observation and theory are needed to characterize the series. "With some practice these few data give the reader such a clear and complete picture of the manner of distribution as cannot be obtained from a word description, however complete."[54]

Pearson has shown by illustrations that numerous statistical series from the fields of meteorology, biology, and anthropology, as well as from the fields of demography and economics, may be presented according to his method. Not only has he treated series which are distributed about a mode with decided asymmetry—such as barometric measurements—but he has also proved that even in series of

[53] Pearson, loc. cit., p. 359.

[54] Duncker, loc. cit., p. 33. The variations of Müllerian glands of hogs are characterized by applying Pearson's method in the following way: M (mean) $= 3.5010$, ϵ (standard deviation) $= 1.6808$, Δ (degree of coincidence between observation and theory, that is measure of the lack of coincidence between the empirical and the computed variation polygon according to the manner of computation proposed by Duncker) $= 1.57\%$, n (number of observations) $= 2000$, formula of curve:

$$y = 473.9 \times \left(1 + \frac{x}{4.2889}\right)^{4.8434} \left(1 - \frac{x}{15.6023}\right)^{17.6189}$$

apparent symmetrical distribution—such as series of measurements of heights—a better agreement between the theory and the observations can be obtained by assuming a certain asymmetry than by basing the series on a symmetrical curve.[55] He also examined series of extreme asymmetry, so-called unilateral series which start with the mode, and presented them as a special case under his generalized law of probability. He used as illustrations of this kind of series: the distribution of the houses of England according to value, a distribution characterized by the fact that the houses of the lowest value are the most frequent, and several botanical series, for instance, the number of blossoms and petals of plants of a definite species in which the lowest values are the most frequent, while greater values occur with decreasing frequency.

Pearson found that the types 1 and 4 occur most frequently; botanical measurements usually correspond to type 1, zoological to type 4.[55a]

[55] As opposed to this, Lexis' opinion may be mentioned "that the proof of even an approximate symmetry is of greater theoretical significance than the accurate presentation of a distribution by an unsymmetrical curve the origin of which cannot be conceived as plainly as that of the normal curve." (Article, "Anthropologie und Anthropometrie" in Handw. d. Staatsw., 2nd ed., Vol. I, p. 397.)

[55a] The fundamental distinctions between the two schools led, respectively, by Pearson and Edgeworth are stated as follows by A. L. Bowley: "In the one, the predominate idea is to find a purely empirical formula to fit the observations, the excellence of the formula being measured by the closeness of the fit and the fewness of the arbitrary constants, and then to assume that the observations can be replaced by the mathematical formula, and the laws of chance applied; in the other it is rather sought to find *a priori* what mathematical law will tend to be obeyed if certain postulates are given as to the genesis of the observed quantities, and to discover those postulates which yield a law that adequately describes the phenomena. The methods and the formulæ of the two schools are intimately connected, and in many cases yield identical results." (Journ. Roy. Stat. Soc., Vol. LXIX, p. 747.)—TRANSLATOR.

CHAPTER III

THE DISPERSION OF SERIES COMPOSED OF MAGNITUDES LIMITED IN A DEFINITE WAY (CONSTITUENTS OF A GREATER TOTALITY)

Series whose items denote the sizes of masses limited in a definite way (constituents of a larger totality) form the second group of series distinguished in the beginning of the book. The series of this group are time, space, qualitative, or quantitative series depending on the criterion according to which the items are selected. For example, to this group belong those series the items of which indicate how many births and deaths have occurred in the single months of the year, or in successive years of a longer period, or how many inhabitants are in the different districts of a country, or how many persons follow the various occupations. The series of the second group—as has already been explained—admit only the computation of the arithmetic average of the items (constituent masses) of the series, but not the computation of other means.

First of all, the average of the series may be supplemented by the statement of the extreme values, the maximum and the minimum. In doing this we may limit ourselves to indicating the highest and the lowest figures occurring in the series. But we may also describe more exactly the items to which these extreme values refer, i. e., the years in which maximum and minimum occurred, or the districts which have the greatest or the smallest number of population, etc.[56]

[56] Öttingen speaks of the "tenacity" of a time series with small differences between the average and the extreme values; in case of

If the statement of the extreme values does not seem to be sufficient, then we may resort to the device, also used with series of quantitative individual observations, of presenting certain classes which are of significance in the given case.

Finally, in order to obtain a uniform expression for the dispersion of the series, the average deviation may be computed as a supplement of the average. This computation is effected by combining the deviations of all the items from the average and the result may be stated either in absolute size or in per cent of the average. Different statistical writers, especially Ad. Wagner[57] and G. von Mayr,[58] have frequently used this method in time series. Such measures of dispersion, however, give a clear picture of the deviations in the series under consideration as well as in series of quantitative individual observations only if the items be distributed symmetrically on both sides of the average.

large differences, however, he speaks of the "sensibility" of the series.

[57] Cf. Gesetzmässigkeit in den scheinbar willkürlichen menschlichen Handlungen, "Comparative Suicide Statistics of Europe" (1864), for instance pp. 88 and 93. In the latter place Wagner has given the absolute size of the average deviation as well as its percentage of the mean.

[58] Cf. Gesetzmässigkeit im Gesellschaftsleben (1877), p. 57 ff. In "Statistik der Bettler und Vaganten im Königreiche Bayern" (1865), v. Mayr has merely added the deviations (of the numbers of poor for 1835-1860) from the average number of poor for this period, and has used this sum as measure of the deviations without computing a real mean deviation (by dividing the sum of the deviations by their number). (Cf. ibid. pp. 19 and 28.) Also, Engel occasionally used indices of fluctuation; thus in his Bewegung der Bevölkerung in Königreich Sachsen in den Jahren 1834-1850 he computed the mean duration of marriage by dividing the number of existing marriages by the number of weddings, correcting the latter by a coefficient computed from the mean annual fluctuation of the frequency of marriages.

The time series consisting of absolute numbers, which form the most important subdivision of the group of series under consideration, however, very frequently show a characteristic conformation which does not lie in the distribution of the items about the mean but in some other regularity—for instance, a direction of development progressing in a definite manner and keeping pace with time, or a definite periodicity. If such a conformation is present, then the measurement of the dispersion of the items about the mean is not sufficient. In such a case the actual characteristic property of the series, for instance, the direction of its development or its periodicity, cannot be seen from the size of the measure of dispersion. In order to ascertain this property, the items must be studied successively and be presented by means of other methods discussed in Appendix I, because when measuring the dispersion we do not take the succession of the items into consideration but merely measure their distances from the average.[59]

The fact that statistical series with only slight time fluctuations have frequently been found, has caused many discussions in statistical literature. On the basis of the observation of the invariability, constancy, or steadiness of certain series, statistical "laws" were constructed, frequently without adequate basis, as in the case of Quetelet's budget of penitentiaries and scaffolds. Likewise, important philosophical conclusions were drawn concern-

[59] If the manner of the development or of the periodicity of a series is already established, we may try to measure independently the irregular individual fluctuations occurring in the series superimposed upon the regular fluctuations, and to compute their average. For this purpose we must ascertain the value for every year which results merely from the general development or periodicity, and in this way we construct, so to speak, a hypothetical curve. The deviation from the hypothetical curve (i. e., from the value resulting in the hypothetical curve for the year in question) of the actual series must then be ascertained for every year separately and the average of these deviations must be computed.

ing free-will, which naturally caused endless controversy. As a matter of fact, there are only few steady time series consisting of absolute numbers. Most phenomena that are treated statistically are closely connected with the number of population. But the latter shows a more or less decided variation in almost all countries. Consequently, a definite trend is usually expressed in the various social phenomena whose absolute sizes are ascertained statistically. Only by computing relative numbers which express the extent of the phenomena observed per thousand of the population or per inhabitant is it possible to obtain values independent of the variation of the number of population. Therefore, the question of the permanence and so-called regularity of the social phenomena can only be solved by investigating the fluctuations shown by series consisting of relative numbers. These series belong in the third group and their dispersion will be treated in the following chapter.

An indubitable connection between the homogeneity of the series and its dispersion exists in the time series formed of absolute numbers, as well as in the series of quantitative individual observations. Thus, a series which extends over periods in which the phenomena measured were exposed to very heterogeneous influences, undoubtedly shows greater variations than a series which refers to a period that in itself is homogeneous. If a time series is decomposed into constituent series the items of each being homogeneous in character, for instance, if economically favorable and unfavorable years are separated when investigating the fluctuations of births and marriages, then constituent series are obtained which usually show much smaller fluctuations than the total series. Adolf Wagner divided the marriages in Belgium during the period 1841-1858 into three groups by combining years of the same industrial character. He distinguished a normal period consisting of the years 1841-1845, an unfavorable period consisting of the years

of dearth, 1846, 1847, 1854, 1855, and a favorable period consisting of the years of low prices, 1849, 1850, 1857, 1858. He found an astonishing regularity for each period.[60]

[60] Cf. Gesetzmässigkeit in den scheinbar willkürlichen menschlichen Handlungen, "Comparative Suicide Statistics," p. 934.

CHAPTER IV

THE DISPERSION OF SERIES OF RELATIVE NUMBERS AND AVERAGES WHICH CHARACTERIZE MASSES LIMITED IN A DEFINITE WAY (CONSTITUENTS OF A GREATER TOTALITY) IN SOME OTHER RESPECT THAN AS REGARDS THEIR MAGNITUDES

A. THE GENERAL PROBLEM (DISTINGUISHING TIME, SPACE, AND QUALITATIVE OR QUANTITATIVE SERIES)

The third group embraces those series of relative numbers and averages which characterize masses limited in a definite way (constituents of a greater totality) in some other respect than as regards their magnitudes. The *relative number* computed directly for the greater totality stands as a mean of the items of the series, if these are relative numbers; if the items of the series are themselves averages of quantitative elements of observation (computed for the constituent masses), then the *average* which is computed for the totality, and which is logically analogous to the items, is to be regarded as the mean of the series.

The dispersion of a *time* series consisting of relative numbers or averages expresses the degree of stability (steadiness, constancy) or the degree of variability of the phenomenon measured. Phenomena which nearly coincide for consecutive periods of time, or which give rise to relative numbers or averages but slightly different from the general averages of a longer period, are "stable" in contradistinction to the "variable" phenomena whose values present decided time fluctuations. Such time series may be

classified, therefore, according to the size of their measures of dispersion. It is, however, to be kept in mind that time series of relative numbers and averages, like time series of absolute numbers,[61] frequently exhibit a characteristic conformation, not in the grouping of the items about the mean but in some other way, such as a definite evolution or a definite periodicity. This characteristic conformation cannot be established by merely measuring the dispersion of the items about the mean, as such a measure takes no account of the order of the items, which is the essential element in determining the evolutionary or periodic character of the series. Therefore, in this connection, chief attention will be given to those series which possess no marked evolution or periodicity, while the investigation of the series which have a characteristic conformation, but not in the grouping of the items about the mean, will be considered in Appendix I.[62]

Time series of relative numbers and averages usually exhibit relatively slight fluctuations in the course of years. If this were not true, then most statistical data would possess no practical value, as conclusions could not be based upon them. Small time fluctuations are to be found, as we should expect, primarily in the fields of anthropology and meteorology. The average height and weight of the inhabitants of a country change very little. Likewise, there is little variation in the average temperature, barometric height, or rainfall of a given country. Also in the fields of social, moral, and economic statistics there exists relatively great stability, as is evidenced by the sex-ratio among births, the average and the normal lengths of life, the

[61] Compare with p. 294 f., above.

[62] It is self-evident that the dispersion of evolutionary or periodic series can also be measured. Such measurements, however, are not concerned with the essential nature of the series and possess significance only in individual cases for some particular purpose.

birth, marriage, and death rates, the percentage of crimes, the consumption of food per capita, the average income, the average wage, and numerous other values which exhibit very slight fluctuations during consecutive years.[62a]

The attitude of statisticians in regard to the significance of the stability of statistical values has changed decidedly during the last few decades. When the relative stability of a large number of social phenomena was first established it was believed that an extremely remarkable discovery had been made. The regularity of births and deaths spelled "göttliche Ordnung" to Süssmilch, while the invariability (Gesetzmässigkeit) of demographic and moral statistical phenomena filled Quetelet and his followers with admiration for the "natural law" that seemed revealed. The constancy observed for a few phenomena, limited periods of time, and a restricted area was represented as being universally true and, where observations were wanting, constancy was presumed without further investigation. The many data collected in the course of time have since shown, however, that a variety of degrees of stability (or variability) exist, and that the degree of variability of a single phenomenon changes in the course of time and from country to country. The stability of social (demographic, moral and economic) phenomena is, therefore, at present not considered a general law, and where constancy is found it is explained without the aid of metaphysics and, as a rule, without postulating natural law. The constancy or variability of a social phenomenon is, according to present opinion, due to the constancy or variability of the conditions upon which the phenomenon in question is dependent. If the complex of causes which produces the phenomenon re-

[62a] The author is in error in stating that the variation in the average temperature, rainfall, and barometric height of a given country is small, as it varies by as much as 100% in certain districts. Likewise, certain series of social statistics, such as annual divorce rates, have wide fluctuations.—TRANSLATOR.

mains, in the main, unchanged, then there is no possibility for an essential change in the phenomenon. If, however, any of the causes is altered then the dependent phenomenon must change.

The phenomena with which statistics is concerned depend both upon natural and upon social causes. The former predominate, of course, in the fields of meteorology and anthropology. In demography the limitations of human life and of the human reproductve period are conditions prescribed by nature. The ratio of the sexes also appears to be ruled by natural law. Probably it is also to be considered a result of natural law that a great part of the children born are lacking in vitality and succumb to a special mortality.[63] The general composition of a population according to sex and age is, therefore, determined by natural law, and it can alter but slowly. The natural stability of the demographic divisions of population necessarily results in relatively great social and economic stability and hence in a certain regularity of social and economic mass-phenomena.[64]

Various social causes are acting in the general framework of society. These causes, such as occupation, economic condition, education, etc., exercise, as is their nature and intensity, a varying degree of influence upon demographic, moral, and economic phenomena. These various degrees of influence are, however, quite compatible with the constancy of the phenomena named. One can start with the idea of dividing the population according to occupation, economic condition, education, etc., into more homogeneous groups which participate to a varying degree in demographic, moral and economic phenomena. The eco-

[63] Lexis, Abhandlungen zur Theorie der Bevölkerungs- und Moralstatistik, V, "Concerning the Causes of the Slight Variability of Statistical Relative Numbers," p. 87.

[64] Ibid. pp. 353 and 94. Also see the articles "Gesetz" and "Moralstatistik" by Lexis in the Handw. der Staatsw.

nomic and intellectual conditions of a definite group of the population, and the hygienic conditions under which the members of that group live, naturally cause a definite mortality; they make it possible for a definite percentage of the young people to marry and to have children; they also expose the members of the group to definite temptations to commit crimes of one kind or another and thus cause a definite rate of criminality. The various groups of the population have also, of course, various economic requirements and produce various economic goods. As long as the conditions of life of the members of the different groups do not change, so long will each group develop the same intensity of various demographic, economic, and moral phenomena. If, at the same time, the whole population continues to be composed of the same constituent homogeneous groups, then, of course, the whole population will continuously exhibit demographic, moral, and economic relative numbers and averages of the same size. This is true —always assuming that the composition of the population remains the same—even if the number of population increases or diminishes, since the sizes of both relative numbers and averages do not necessarily vary with the number of observations. The death rate (number of deaths per thousand living) or the per capita consumption of meat may have the same values for wholly different numbers of population. Of course, if the composition of the population changes in any manner, that is, if the proportion existing among the various constituent homogeneous groups changes, then an effect will always be noticed in some one or other of the phenomena comprehended by statistics. Thus, for example, an increase of industrial activity, which reduces the numbers of the unemployed and places many people in a better economic position, will diminish the mortality and the number of crimes against property, but will increase the number of marriages; an economic crisis, on the other hand, swells the weakest economic classes, and these classes

will contribute more strongly, according to their peculiarities of conduct, to the various demographic and moral phenomena.

From what has been said it follows that general laws of stability in the province of social statistics cannot be established. The relatively great stability of social phenomena can be explained by the permanence of the underlying cause-complexes; that is, the stability naturally results since the economic institutions and mental make-up of a given population change very gradually, while statistical series usually refer to relatively short periods of time. This stability has, however, nothing absolute in it, but it is merely an empirical fact of observation arising from definite concrete conditions and it may at any time give way either to marked fluctuations or enter a definite evolution following some political, economic, technical, or other cause. If we desire to use the term "statistical law" in its broad sense in referring to the stable phenomena that we have been considering, we must keep in mind that we are dealing with purely empirical social laws, which are to be considered merely as historical categories.[65]

In statistical literature attempts have been made to divide time series into groups of a definite character, which groups would also exhibit various degrees of stability. These attempts have, as will be shown, led to no satisfactory classification, as essentially related series often possess quite

[65] According to Wundt it is not correct to speak of an empirical "law" with reference to the time-constancy of a phenomenon, as there is, in this case, no pertinent characteristic indicating a causal relationship. On the contrary, those regularities are to be designated as empirical laws in which the mass-phenomena stand in a functional relation to definite time- or space-values (for instance, a regular evolutionary tendency), or in which there is directly concerned a causal relation between mass-phenomena independent of one another (for instance, in the proof of the various death rates in different occupations). (See Logik, Vol. II, Pt. II, Logik der Geisteswissenschaften (1895), pp. 144, 464, 472 f.)

different dispersion, while, on the other hand, the dispersion of quite unlike series is, many times, the same.

Thus the theory has been advanced that statistical phenomena exhibit varying degrees of stability according to the predominance of natural events or of acts of the human will in causing them. It is very significant that opposite views are held by eminent writers; some holding, on the one hand, that phenomena that depend upon natural factors are more stable and others holding, on the other hand, that it is the actions controlled by the will, such as marriage, criminality, and suicide, that recur with greater regularity.

In general, it is certainly true that natural causes bring about fluctuations that are often between narrow limits. The sex-ratio of births, which seems to be determined exclusively by natural causes, is the most stable demographic phenomenon which we know. Likewise, the average stature of the population of any country scarcely varies, although, it is true, this value is also influenced by social causes which help or hinder growth. On the other hand, other phenomena, also chiefly dependent upon natural causes, show greater fluctuations. For instance, meteorological phenomena (average rainfall, average temperature, average barometric height, etc.) do not only vary widely with the seasons but also from year to year. Consequently, agricultural products, which are dependent upon meteorological conditions, likewise show considerable fluctuations. It is quite different, however, with certain phenomena which are dependent chiefly upon the human will, such as marriages, crimes and suicides, which show a remarkable regularity during the course of years. The comparison of these "voluntary" phenomena with mortality, which is subject to natural laws, is the chief origin of the thesis of the greater regularity of voluntary events. It is true that there are greater fluctuations in mortality rates than in marriage,

crime, or suicide rates. But the death rate is a phenomenon by no means exclusively determined by natural causes. The extreme limit and, perhaps, the normal length of men's lives are fixed by natural law. The normal lifetime, as a matter of fact, is extraordinarily stable. The frequency of deaths, however, which does not arise entirely from the normal mortality but largely depends upon the mortality of children, is in a great measure dependent upon economic relations and fluctuates with each economic change. A business depression throws a large number of men out of work or at least forces them to adopt modes of life much inferior and, therefore, dangerous to health; an improvement of the economic situation also influences mortality by enabling many people to adopt a mode of life more advantageous to health. The number of marriages, as well as crimes and suicides, is, to be sure, also influenced by economic conditions, but, in general, to a less degree than is mortality. Marriages, crimes, and suicides are events which are not—like deaths—independent of the human will, but they are conscious and, as a rule, maturely considered human acts. The motives which lead to marriages, crime, or suicide are not exclusively economic; that is to say, these phenomena are, to a great extent, caused by motives which are independent of the fluctuations of economic activity. In consequence of the relative permanence of the general social relations and of the mental constitution of the population, the non-economic motives act with nearly the same intensity during a course of years and sweep nearly the same proportions of the population to marriage, crime, and suicide year in and year out—of course with the difference that each year other individuals are vehicles of the motives—and thus a surprising regularity of "voluntary" action results. In this sense it is undoubtedly certain that human will is, in general, an element of stability. A general proposition that "voluntary" actions are

always more stable than occurrences in which natural causes predominate cannot, however, be established.[66]

The fact of the relatively great stability of crimes, suicides, and similar moral-statistical phenomena has led to many philosophical controversies concerning the freedom of the will. A number of the earlier statistical writers interpreted the regularity of moral-statistical phenomena as " natural law " and believed the freedom of the will to be denied by this regularity, or at least to be confined between definite limits. Quetelet initiated this line of argument. He put forth the idea of natural law for mass-phenomena, without, however, excluding individual freedom of will.[67] This manifestly unsatisfactory formulation was accepted by numerous statisticians, especially by the Italian statisticians Messedaglia, Corradi, Bodio, and Morpurgo, without solution of its inherent contradiction.[68] Other writers, influenced by Quetelet, especially Buckle, went even further than he and used statistical regularity as an argument for a thoroughgoing determinism. Strong opposition to this development arose among the champions of the freedom of the will, especially among the Germans Wappäus, Rümelin, Rheinisch, and others, who would admit no subordination of individual action to the " laws " ruling

[66] This conclusion also results from consideration of statistical series from the standpoint of the theory of probability. Such consideration shows that the laws of chance are independent of the causes which are operative in individual cases. (See the review of Lexis' Abhandlungen zur Theorie der Bevölkerungs- und Moralstatistik by v. Bortkiewicz in the Jahrb. f. Nat. und Stat., 3rd series, Vol. XXVII, 1904, p. 253.)

[67] Quetelet held, moreover, that this moral-statistical regularity was not entirely invariable. He recognized, for instance, that the progress of civilization must carry with it a decrease of mortality and of criminality. But he believed in the possibility of very gradual changes only. (See Über den Menschen, German edition, 1838, pp. 10 f., 557 f.)

[68] See Meitzen, Geschichte, Theorie und Technik der Statistik, 2nd ed., p. 61.

masses, but they offered no satisfactory explanation of the consistency of freedom of the will and social regularity. Adolf Wagner called attention in 1864 to the contradiction between the regularity of mass-phenomena, which he demonstrated statistically, and the freedom of the will, to which he declared his adherence, and said that the removal of this contradiction must be the purpose of further scientific research.[68a] It was Drobisch's reduction of the acts of individuals to the volition resulting from definite motives that first gave the correct standpoint for explaining the relation between the will of the individual and the actions of the many.

The modern explanation based upon the writings of Drobisch, Schmoller, and Knapp, and formulated most precisely by Lexis, ascribes the regularity of moral-statistical phenomena to the relative permanence of the social conditions which are the fundamental causes of such phenomena, the permanency of these conditions regularly giving rise to motives leading to the same acts. This explanation is independent of the metaphysical question of the existence or non-existence of free-will, and it is, therefore, the general conviction that statistics has not to solve and cannot solve this question.[68b]

The extensive moral-statistical material now available shows also that the regularity assumed by the earlier statisticians is usually not so great as the analogy with natural law would make necessary. The mathematical statisticians in particular have shown that even the most stable series of moral statistics which have been observed are affected by fluctuations that are at least as great, and

[68a] Gesetzmässigkeit in den scheinbar willkürlichen menschlichen Handlungen, p. 79.

[68b] Mr. F. H. Hankins has clearly explained the attitude of modern statisticians toward this question in his Adolphe Quetelet as Statistician (Columbia University Studies in History, etc., Vol. XXXI, No. 4).—TRANSLATOR.

usually much greater, than the accidental errors of the values obtained empirically in games of chance. From this it follows that even in such stable series there are disturbing factors whose effects are as great as those of accidental causes. Free-will can enter as one of these disturbing factors operating in a manner analogous to accidental causes. The actual fluctuations of series of moral statistics offer, therefore, sufficient scope for the effects of free-will, understood in the sense of an accidental cause.[69]

The second criterion that writers have believed would divide time series—and, in this case, merely time series consisting of relative numbers—into two groups of different degrees of stability is the differentiation of such series into, first, series of relative numbers which represent the composition of definite masses and, second, series of relative numbers which give the frequency or intensity of definite phenomena. Relative numbers of the first kind are, as a rule, subordinate numbers which bear the character of relative (analytical, secondary) probability, as, for example, the percentage of men or women in a whole population. Relative numbers of this kind, however, may sometimes be coordinate numbers—as the number of women per one hundred men—which correspond to a known function of a relative probability. Relative numbers of the second kind, which give the frequency or intensity of a phenomenon, are always coordinate numbers which can bear the character of a " genetic " probability. In this class belong the mortality rate and the probability of death, the birth rate, the marriage rate, etc.

It has been observed that series of relative numbers which represent the composition of definite masses exhibit a much more remarkable stability in time, that is, the dispersion is less, than do the coordinate numbers which show the fre-

[69] Cf. Tschuprow, "Die Aufgaben der Theorie der Statistik," Schmoller's Jahrbuch, 29 Jahrgang (1905), p. 461 f., and Westergaard, Die Grundzüge der Theorie der Statistik, p. 282.

quency or intensity for the same space of time of the phenomenon whose inner composition is represented by the former series. Most striking are the different degrees of stability shown by a comparison of the birth rates (frequency or intensity) and the sex-ratios of births (inner composition). The birth rates may show very considerable fluctuations while the sex-ratios remain constant. This fact, however, evidently depends upon the further fact that the sex composition of births is exclusively or almost exclusively determined by natural law, while birth rates depend upon various fluctuating social factors. Also the fact that the percentage of stillborn among all births scarcely changes, in spite of the fluctuations of the birth rates, is probably most intimately related to the natural causes which bring about still-births. But we also discover similar conditions in phenomena whose structure is in no way primarily determined by natural causes. For example, if those marrying be classified according to age and conjugal condition the classes will, as a rule, be much more stable than the general marriage rate. Likewise, the classes of criminals according to age and sex fluctuate much less than general criminality, and suicides classified according to age, sex, and the particular form of death are less variable than the total number of suicides. The number of emigrants per thousand of population varies extraordinarily, nevertheless the age and sex composition remains about the same from year to year. The regularity in the proportion of various classes of mail, in spite of the increase in the volume of mail matter, is well known. The percentage of letters registered, of letters free of postage, of postcards fluctuates very little; likewise the percentage of dead letters remains almost constant.

This great constancy of certain subordinate numbers is, however, not difficult to explain. The explanation is that the factors which cause the variations in frequency (intensity) of the respective phenomena exercise no, or but

slight, influence on the distribution of the events into the various subdivisions considered. Thus, economic factors affect the number of marriages and cause the marriage rate to fluctuate. But these economic factors (economic expansion, crises, etc.) operate more or less uniformly on all marriageable persons regardless of their social position or age, and the subdivisions of the people marrying according to social position or age remain relatively the same, regardless of the changes in the frequency of marriage. The percentage of registered letters expresses the view of the public as to the reliability of the postal service and the risk of sending an ordinary letter. If this view remains the same the percentage of registered letters will remain constant even though the total number of letters is greater.

No generalization can be drawn, however, concerning this group of cases, in which the inner composition is probably more stable than the frequency or intensity of the phenomenon. The question whether the frequency of a phenomenon fluctuates more than do the classes obtained by dividing the individual items into definite categories, must be solved *de novo* for each phenomenon. Finally, it also happens that where the classes exhibit extraordinary stability, changes appear which must not be overlooked. Thus, at present a tendency is shown in most countries, in consequence of the decrease of mortality and the development of industry, for the number of first marriages of both sexes to increase at the expense of second or later marriages. In the statistics of suicides, classified according to the various methods of death, absolute constancy is not possible on account of the changing classification; our modern life gives rise to new and formerly unknown methods, such as death from a moving train. The classes of deaths according to cause in general show a remarkable constancy. In spite of this, certain causes of death fluctuate considerably or show a definite tendency of development. Thus, in England during the period 1871-1890

there was a decrease of mortality from consumption, small-pox, and typhoid and related fevers, while the mortality from cancer greatly increased.[70] Statistical masses in which the inner composition changes more in the course of time than the intensity are, consequently, not unknown. Thus Öttingen has given statistical data concerning the literary productivity of Gemany,[70a] from which it appears that during the years 1850-1875, when the annual production was tolerably constant, the relative proportions of single classes of writing altered considerably. There was an increase of technical books relating to industry and a decrease, especially, of theological and religious literature, whose place was taken by pedagogical books of a popular or juvenile character.

For series consisting of relative numbers or averages, just as for those consisting of absolute numbers, there is a *connection between the homogeneity of the series and its dispersion.* If a time series covers a homogeneous period it naturally exhibits smaller fluctuations than one covering a period during which the factors chiefly influencing the phenomenon represented have varied in strength. If one separates the economically good and bad periods from each other, then less dispersion is shown by those series dependent upon economic conditions (mortality, crimes, suicides, etc.) than by combining years of business depression and expansion.

Aside from economic conditions other causal factors may also be concerned. Thus, the meteorological conditions of a year doubtless exercise a great influence upon the mortality of that year. In general, we find that years with hot summers and cold winters have a high mortality and, inversely, that a cool summer and mild winter are healthful. If one should select from a period of years those which have approximately the same conditions as regards

[70] Compare G. v. Mayr, Bevölkerungsstatistik, p. 323.
[70a] Moralstatistik, p. 555 f.

temperature, and should study the fluctuations of the mortality during these years, one would find a narrower range of deviations from the average than is shown by all the years of the period regardless of temperature.[71]

We now leave time series to consider *geographical series* of relative numbers or averages. The dispersion of such series gives a standard for measuring the variability of phenomena with reference to space. There may be values for various countries or their subdivisions which differ very little from each other and from their average; on the other hand, values for different geographical divisions may differ widely. In general, geographical series, as long as the items do not refer to divisions of a single homogeneous country, exhibit greater fluctuations than do time series. This is very easily understood. That time series made up of demographic, moral, or economic statistics exhibit but small fluctuations rests upon the fact that, as a rule, only a few decades are considered, during which social conditions usually change very little. On the contrary, it is to be expected that values drawn from different countries will usually show greater differences for different countries develop along a great variety of lines because of differences of geographical position, climate, or other natural attributes. In addition there are often ethnological differences of population, and also differences of political, economic, and psychological development, which are the results, not of tens of years, but of hundreds of years of history. For these reasons there are essential differences of social and demographic conditions among the populations of different lands, which are reflected to a greater or less degree in all statistical phenomena. They appear, however, to the smallest degree in those phenomena which seem to rest upon definite natural laws rather than upon social factors. Thus, there is an extraordinary degree of correspondence among the sex-ratios at birth in various

[71] Cf. Westergaard, Die Grundzüge, etc., p. 19.

countries.[72] Greater differences are to be found in the mean and normal length of life of various populations. That meteorological and anthropometric averages (rainfall, temperature, barometric height, stature and weight of the inhabitants, etc.) vary decidedly for different countries is well known. Likewise, the differences of most demographic and economic relative numbers and averages are considerable, but not such as to enable us to make any general propositions concerning the dispersion of geographic series. Various phenomena (birth rates, mortality, the per capita meat consumption, etc.) as a rule show various degrees of dispersion, and it is, therefore, necessary to depend upon actual observation to determine the kind and magnitude of the geographical differences of each phenomenon.

Because of the fundamental differences which, as a rule, exist between various countries as regards the composition and activity of their populations, we usually do not compute general averages for several countries together, but limit ourselves to stating the minimum and maximum which express the range within which the values for all countries appear. The computation of an average embracing all of the countries in the series and the use of it as a basis of comparison appears to be allowable only when the various countries are more or less similar in nature; for example, countries which possess the same type

[72] In almost all countries (according to Bodio's compilation based for the greater part upon the years 1887-1901) there are 104-106 boys to 100 girls among living births. England, with a ratio of 103.4, is the only country under 104, while only Spain, Portugal, Greece, Roumania, and Connecticut have over 106. It is possible that in a geographical comparison the differences in the ratio are due to racial differences. Edgeworth has ascribed the exceptionally large excess of boys in Wales to the Celtic descent of the inhabitants. (Journ. of the Roy. Stat. Soc., 1898, p. 130 f.) It is known that there is a greater excess of boys among the Jews, which may depend, however, according to G. v. Mayr (Bevölkerungsstatistik, p. 188), upon the fact that crossing is unusual and inbreeding more common.

of civilization, as those of middle Europe, or where the items of these series do not relate to different countries but to parts of a single country, within which we may expect approximate coincidence. Thus, it is customary to compare the density of population of sections of a nation with the national average, that is, to ascertain in what manner the relative numbers which express the density of a district are distributed about the relative number (found by dividing the total population by the total area) which expresses the density of population of the entire nation. In the same way, it can be found out in what manner the average income of the residents, the average age of those marrying, etc., of the several districts are grouped about corresponding national averages.

The proposition concerning the connection of the homogeneity of series and their dispersion holds for geographical series of relative numbers or averages as well as for the corresponding time series. That is, less dispersion is obtained by decomposing such series into homogeneous subdivisions. For example, if one differentiates nations as healthful or unhealthful according to their various climatic or meteorological characteristics, then, in all probability, the series, such as death rates, relating to each homogeneous group of nations will show much less dispersion, that is, less deviation from the average, than the series relating to all nations regardless of climatic or meteorological characteristics.

Finally, *quantitative and qualitative series of relative numbers or averages* will be discussed with reference to their dispersion, in order to get a picture of the grouping of the items about their average. We compute the death rates in various occupations and determine between what limits and in what manner the various death rates group themselves about the average death rate of the whole population. Likewise, a series may be composed of the average wages of different occupations and the group-

ing of these about the general average may be determined.

But there are reasons, peculiar to the nature of qualitative and quantitative series, that make the investigation of the dispersion of such series of less significance than that of geographical series and time series. If a phenomenon (such as mortality) remains constant or fluctuates but little during a term of years, we could not have predicted it as a matter of course. It is only by study of the actual figures that the dispersion is disclosed and a measure of the variability of the phenomenon with reference to time is obtained. Likewise, it cannot be predicted whether a definite phenomenon will appear uniformly or with varying intensity in different geographical districts; investigation of the dispersion of the geographical series in question must give the information.

It is quite different with qualitative and quantitative series. Such series, as a rule, are built upon the basis of some mark of differentiation which we already know or, at least, assume to possess causal significance for the phenomenon represented by the series. Thus, we divide those dying according to occupation or economic condition, because we know that these influence mortality and that the members of different occupations and economic classes exhibit different death rates. We therefore expect differences in the items of qualitative and quantitative series.

However, it is not so much the deviations of the individual items from the average that interests us as it is their interrelations, the comparison of their relative sizes. Thus, the grouping of the death rates about the average, if deaths be classified by occupation or economic position, is not very significant, but it is important to ascertain the relative mortality of different occupations and economic classes, and to determine the amount of difference between the individual items. Only in this way can we find out, for instance, if mortality according to well-being exhibits

a definite characteristic conformation, other than a special grouping of the items about the average, the mortality perhaps varying inversely with economic well-being. In addition to these more theoretical objections, the measurement of the dispersion of qualitative and quantitative series is of little use, because such series usually consist of a small number of items whose relationships can be comprehended without special computations.

The reason that qualitative and quantitative series usually exhibit greater fluctuations than geographical and time series is, therefore, that the former, as a rule, originate from the use of a criterion of a causal character. If, for example, economic position influences mortality, then it is obvious that greater differences of mortality must exist between the poor and the rich of a country than between the total death rates of consecutive years during which the economic organization changes but little, or between the total death rates of different countries whose average economic positions are not usually as extreme as those of the lowest and highest classes of the same country. It is self-evident that as natural causes operate to bring about time stability and geographical coincidence, so also to that extent they tend to produce greater coincidence of qualitative and quantitative constituent masses. To illustrate, as the normal stature of a population appears, so to say, to be predetermined by nature, the various social classes, of course, exhibit but slight differences of stature. The influence of social factors is greater upon the normal length of life, which is largely, nevertheless, dependent upon natural causes. Likewise, sex-ratios of children born, when found for qualitative and quantitative constituent masses, in spite of the natural causes at the basis, exhibit not unimportant variations in contrast to the great stability in time and in geographical distribution. The percentage of boys is known to be much greater among stillborn than among living births; on the contrary, it is less among

illegitimate births than among legitimate ones and, at least in Germany, less in large cities than in the country.[73]

B. ELEMENTARY MATHEMATICAL METHODS FOR MEASURING AND REPRESENTING THE DISPERSION OF SERIES OF RELATIVE NUMBERS AND AVERAGES WHICH CHARACTERIZE MASSES LIMITED IN A DEFINITE WAY (CONSTITUENTS OF A GREATER TOTALITY) IN SOME OTHER RESPECT THAN ACCORDING TO MAGNITUDE

The elementary mathematical methods by means of which the dispersion of series belonging to the third group (whether they be time, space, qualitative, or quantitative series) can be measured and represented are, for the greater part, the same as those used for series consisting of individual observations, or of items which give the magnitudes of definite masses (first and second groups). The simplest method consists in presenting the extreme members, maximum and minimum, of the series together with the average. These values give the range within which lie all the items of the series. In addition, the location of the extreme values may be given. Thus, the dispersion of the death rates for a series of years, for the provinces of a country, or for different occupations, may be character-

[73] Lexis has explained the slighter excess of boys among illegitimate births and in large cities by suggesting that such classes may be distinguished by different percentages of premature births and has thus attempted to harmonize these phenomena with the regulation of the sex rates by natural law. (Cf. Abhandlungen zur Theorie der Bevölkerungs- und Moralstatistik, VII, "The Sex-ratio of Births and the Calculus of Probability," pp. 166-168.) G. v. Mayr explains the greater excess of boys in the country (as with Jews) by the fact that crossing is more unusual and inbreeding more common.

The question whether the age relation of the parents, in the sense of the Hofacker-Sadler hypothesis, has an influence upon the sex-ratio of offspring is, as yet, undecided; so are the questions of the influence of fecundity and nourishment.

ized by the average death rate and the extreme values, the year, province, or occupation to which the latter refer being denoted as well as the numerical size of these values.

The presentation of more than one average, by which we can suitably characterize the dispersion of series of the first group consisting of individual observations, is impossible for series of the third group and unusual for those of the second group. If the presentation of the extreme values does not appear to be sufficient the average may be supplemented further—as in the case of series of the first and second groups—by the presentation of certain classes which possess especial significance

Under certain circumstances we may also consider the computation of the average deviation, which may be given by its absolute size or as a percentage of the average of the series. Such a computation, however, is especially difficult for series of relative numbers or averages. The items of such series possess different weights, and the weights are not indicated by the series. If we take the deviation of each member of a series of relative numbers or averages from the average of the series and then compute the simple arithmetic mean of these deviations, we assume that the various members have equal weights, which is usually contrary to fact. The assumption of the equality of weight of the members of time series usually introduces but little error. For example, if the population of a country has changed but little during the years considered, which is often the case for a short period, then the yearly death or birth rates may be considered of equal weights and the average deviation of death or birth rates for the whole period may be computed.[74]

[74] Thus Adolf Wagner in his Gesetzmässigkeit in den scheinbar willkürlichen Handlungen (p. 88) has computed the mean numerical deviation of marriage and birth rates in order to determine the time-movement of these items for various countries, and Mayo-Smith in his Statistics and Sociology (p. 91) has done the same for Ger-

In geographical and qualitative and quantitative series, however, the items are, as a rule, of such decidedly different weights that the computation of the simple arithmetic average of deviations is not allowable. Imagine a series of death rates for various provinces or for various occupations. The individual death rates are, in all probability, of different weights, as the provinces are not equal in size, nor do the occupations contain equal numbers. The computation of the simple arithmetic average of the deviations, by which it is assumed that the death rates have the same weights, would give an incorrect result. For example, if the larger provinces or occupations vary but little, while the smaller provinces or occupations vary widely from the average, the average deviation would be greater than the facts justify. In order to correctly compute the mean deviation one would have to take into consideration the various weights of the deviations of the series, that is, proportional weights would have to be assigned to the members and a weighted average computed. However, such a computation would, as a rule, require work out of all proportion to the value of the measure of fluctuation desired.

man birth rates. Both authors give the mean deviation as a percentage of the average of the series. The figures for the different months of a year may be considered of equal weight, if we are not striving for great accuracy, and their dispersion characterized by an index of fluctuation. Thus, Kollmann in his Die Bewegung der Bevölkerung in den Jahren 1871 bis 1887 mit Rückblicken auf die ältere Zeit has used indices of dispersion, computed in the following way, to sum up the various degrees of the monthly variation of marriages, births, and deaths in the Grand Duchy Oldenburg: He computed the daily average for each month and for the whole year; assuming the latter to be 1,000 he expressed each average for a month proportionately; finally, he found the deviations of the relative numbers for each month from 1,000 and averaged them. (Cf. Statistische Nachrichten über das Grossherzogtum Oldenburg, No. 22, 1890, pp. 25, 83, 114.)

C. EXAMINATION OF THE DISPERSION OF SERIES OF NUMERICAL PROBABILITIES BY THE METHODS OF THE THEORY OF PROBABILITY

The theory of probability offers a special method for examination of the dispersion of series whose items are in the form of numerical probabilities or their functions.[75-76] That is to say, such series may, with reference to their dispersion, be compared with series which originate from some game of chance, for example, the drawings of balls from an urn in which there are a definite number of red and white ones. The analogy between statistical probabilities and the results of games of chance already set forth (p. 171 f.), which is assumed when the theory of probability is applied to statistical material, may be summarized

[75] Cf. the definition of numerical probabilities, p. 19 f.

[76] Relative numbers which do not possess the character of numerical probabilities (or their functions) may sometimes be treated according to the theory of errors of observation like the absolute numbers originating from repeated observations of the same object. It can accordingly be ascertained whether the items exhibit a grouping corresponding to the law of error; if this be the case, then the items may be considered to be accidental modifications of a fixed "typical" or normal value and the dispersion may be expressed in conformity with the theory of error (the "physical" method) by means of the mean or probable error or some other proper measure. Cephalic indices (which are relative numbers obtained by taking the ratio of the breadth of the cranium to its length), especially, have been treated in this way and the dispersion has been proven to correspond to the normal law of error. We can thus, perhaps, assume that the average cephalic index of a definite people possesses a typical value. Likewise a generalization of the law of error may be applied to relative numbers. Thus, Pearson has applied his method of the "generalized probability curve" to a series of Bavarian cephalic indices and established an unsymmetrical grouping which, however, corresponds to an extended law of error. Pearson has also treated English statistics of the percent of paupers by means of the "generalized probability curve." (Cf. "Contributions to the Mathematical Theory of Evolution," II, "Skew Variation in Homogeneous Material," pp. 388, 404.).

with especial reference to the problem of the measurement of dispersion in the following way:

If one ball is drawn from an urn containing red and white balls in the ratio 6:4 the objective (theoretical) probability that a red one will appear is 0.6, and that a white one will appear is 0.4. If the experiment is made of drawing a ball 1,000 times, replacing it after each drawing, the percentage of red or white balls drawn may not exactly coincide with the objective probability (which would be the case if 600 red and 400 white balls be drawn) but will rather closely approximate it. Repeating the experiment and making note of the percentage of red and of white balls in each drawing of 1,000, a series of empirical probability values (that is, empirical expressions for the objective probability of the appearance of a red or white ball, merely affected by accidental errors) will be obtained which group themselves in a characteristic manner about the known objective (theoretical) probability. The peculiarity of the grouping consists in the concentration of the empirical values about the middle of the series and a diminution in the number of values as their deviation from the objective probability increases. The limits between which the empirical values fluctuate depends essentially, in conformity to the "law of great numbers," upon the number of underlying observations. The greater the number of observations, upon which the empirical values are based (called the basic number), the greater will be the precision of the empirical values, that is, so much closer will each empirical value (with a given probability) approximate the objective probability; the values will be more closely grouped, and the dispersion will be less. The smaller the basic number of observations the more will the empirical values deviate from the objective probability and from each other. The amount of dispersion of each of several series of empirical values for the same objective probability depends, therefore, upon the

basic number of observations underlying the individual items. From the objective probability and the basic number, the dispersion to be expected on the basis of the law of chance may be directly computed for any given case. In particular, the standard, the probable, and the arithmetic average deviations of the empirical values from the objective probability (or, as the case may be, from the arithmetic mean of the empirical values which is taken as the most probable value of the objective probability), as well as the modulus and the precision to be expected, can be numerically determined. The dispersion actually originating from a correctly conducted game of chance will approximately coincide with that theoretically determined by computation.

In statistics there is never a known objective probability for the empirical values. The problem which statistics has primarily to solve is whether definite values obtained by observation (relative numbers in the form of probabilities or their functions) may or may not be considered, with reference to the dispersion about their arithmetic mean, to be empirical values of some objective probability. The objective probability must be ascertained from the observations. If the observed values arrange themselves about their average in the same manner and between the same limits, as would empirical probabilities based on the same number of observations originating from a correctly arranged game of chance, then one may conclude that the statistical observations are the consequence of some probability common to them, merely affected by accidental errors. This conclusion is naturally of very great significance. Assuming the conclusion, the observations lose any scientific significance; the important thing being the theoretical probability to which the observations point. The arithmetic average of the observed values can be regarded as the most probable value of the theoretical probability. It may be designated as a "typical" mean with reference to the grouping of the items about it, and it possesses greater

significance than any one of the empirical items. The existence of a theoretical probability signifies that the phenomenon in question arises from natural law or in case it concerns a social phenomenon—for instance, in its time fluctuations—that the general conditions, upon which the phenomenon depends, remain constant, even if disturbing accidental causes bring about small fluctuations so that the stability of the series is not complete. The accidental causes are the equivalent of the contingencies which arise in drawing balls from an urn (among these are the manner in which the urn is shaken, the positions consequently occupied by the balls of different color, the blind selection of the ball, etc.); the constancy of the general conditions, which is indicated by the stability of the statistical series, corresponds to the constancy of the ratio between the balls of the two colors.

If the distribution of a series of relative numbers, which correspond to the formal conditions, coincides with the grouping defined by the theory of probability, then, with Lexis, we may designate the dispersion, or the stability, of this series as "normal." If the relative numbers deviate from each other more than would be the case with a corresponding game of chance, but nevertheless arrange themselves symmetrically about their average like accidental fluctuations, then the dispersion of the series may be called "supra-normal," and the stability "infra-normal"; if the relative numbers deviate less from each other than do the items of the corresponding accidental series then the dispersion of the series is "infra-normal," its stability is "supra-normal."[77]

[77] Cf. Lexis, Theorie der Massenerscheinungen, p. 34, and article "Gesetz" in the Handw. d. Staatsw. The expressions "normal" dispersion and "normal" stability must, naturally, not be understood to mean that such dispersion and stability are the rule in statistics. Such distribution is to be normally expected in the domain of games of chance, the peculiar domain of the theory of probability, but it is a rare exception in the field of statistics.

The methods by means of which one may determine whether the dispersion of a statistical series of relative numbers of the correct form coincides with the dispersion which would result from a game of chance conforming to the theory of probability were developed mainly by Lexis.[78] Westergaard, von Bortkiewicz, Czuber, and Blaschke have also worked in this field. The object, particularly, is to determine the values which would be expected according to the theory of probability and to compare with them the grouping of the values obtained by observation. Instead of comparing the entire distribution of the values obtained by observation and by theory, a simpler method originated by Lexis may be applied. This method consists, first, in computing the probable error, which gives theoretically (that is, in conformity to the theory of probability) a constant probability for the empirical values resting upon corresponding numbers of observations (computation of the probable error according to the "combinational," "statistical," or "indirect" method), and second, in determining the probable error in the sense of the theory of errors of observation directly from the observed values, whereby the items are considered simply to be consequences of some measurement subject to accidental errors, without reference to the question whether these values may or may not be considered as numerical probabilities (determination of the probable error according to the "physical" or "direct" method). If the probable errors (or the medial or average errors, moduli, or precisions, as one may select) computed by the two methods coincide, that is, if their

[78] Cf. especially "Das Geschlechtsverhältnis der Geborenen und die Wahrscheinlichkeitsrechnung" and "Über die Theorie der Stabilität statistischer Reihen" (which first appeared in the Hildebrand-Conrad Jahrbücher, 1876 and 1879, but are now contained in Abhandlungen zur Theorie der Bevölkerungs- und Moralstatistik) as well as Zur Theorie der Massenerscheinungen in der menschlichen Gesellschaft (1877) and the article "Geschlechtsverhältnis der Geborenen und Gestorbenen" in the Handw. d. Staatsw.

quotient[78a] approximates 1, this shows that the deviations of the items from the mean are of the same kind as those resulting from a game of chance constructed to correspond and that these items may actually be considered empirical values of some constant probability. If the probable error computed according to the physical method is greater than that given by the combinational method then the dispersion of the series is supra-normal, if less, it is infra-normal.[79-80]

As a matter of fact, series of statistical relative numbers which may be considered numerical probabilities or their functions and which exhibit normal dispersion corresponding to the theory of probability occur but rarely. The best known illustration is the sex-ratio of children born, which Lexis has demonstrated, upon the basis of Prussian, English, and French statistics, to be subject to no greater variation both in its space and time distribution—at least for the same country and within a moderate space of time—than is to be expected from the assumption of a constant

[78a] Designated "divergency-coefficient" by Dormoy and "error-relation" by v. Bortkiewicz.

[79] For a more inclusive treatment see Lexis, Abhandlungen zur Theorie der Bevölkerungs- und Moralstatistik, VII, "The Sex-ratio of Children Born and the Theory of Probability," p. 153 ff.

[80] Series of averages may be treated according to the methods of the theory of probability in the same way as series of relative numbers. The essential question is, Can we consider these averages, with reference to their dispersion, as accidental modifications of some base value? If a comparison of the actual with the expected precision shows the existence of a supra-normal dispersion, then it follows that the values under investigation are subject to a time or place fluctuation. Anthropometric and meteorological averages, especially, are investigated in such a manner. (Cf. L. v. Bortkiewicz, "Kritische Betrachtungen zur theoretischen Statistik," Jahrb. f. Nat. u. Stat., 3rd series, Vol. X (1895), p. 342 f., and "Anwendungen der Wahrscheinlichkeitsrechnung auf Statistik" in the Enzyklopädie der mathematischen Wissenschaften, p. 835 f., as well as Czuber, Die Wahrscheinlichkeitsrechnung und ihre Anwendung auf Fehlerausgleichung, Statistik und Lebensversicherung, p. 341 f.)

probability for the birth of a boy or girl.[81] Differences in the degree of probability exist, however, for certain qualitative and quantitative subdivisions. Thus, we find that in most countries there is a smaller surplus of boys among illegitimate births than among legitimate births, and a greater surplus of boys among the stillborn than among the living-born.[82]

But the special sex-ratios of the illegitimate and the stillborn give time series of normal dispersion, just as the sex-ratio based upon all births does and, therefore, possess independent typical character. That there is a constant total probability for the birth of a girl or a boy (with no differentiation of legitimate births from illegitimate, or still-births from living), in spite of the above considerations, is due to the fact that the percentage of illegitimate births or stillborn fluctuates but little. Aside from the two classes named there are probably other qualitative or

[81] Cf. Geissler, "Beiträge zur Frage des Geschlechtsverhältnisses der Geborenen" (Zeitschr. d. kgl. sächs. stat. Bur, XXXV, 1889, Nos. 1 and 2), and Lehr, "Zur Frage der Wahrscheinlichkeit von weiblichen Geburten und von Totgeburten" (Zeitschr. f. d. ges. Staatsw., XLV, 1889, pp. 172 ff. and 524 ff.), and Stieda, Das Sexualverhältnis der Geborenen (1875). For Austria, Czuber (Wahrscheinlichkeitsrechnung, pp. 325-328) has shown that a moderate supranormal dispersion holds for the relative frequency of a male birth among living births for the period 1866-1897. Czuber does not ascribe this fact to a time change of fundamental conditions but to the heterogeneity due to the contribution of many races to the data used. In the less extensive categories of legitimate and illegitimate still-births the dispersion is approximately normal. In England and Wales the fluctuation of the sex-ratio appears to be continually affected by a definite evolutionary tendency; according to the Report of the Registrar General for the year 1893 (p. xxviii) the excess of boys has gradually decreased from 105.4 and 105.0 in the years 1844 and 1845 respectively, to below 104 in the year 1893, and it has not been as high as 104 0 since 1885. Upon the influence of race upon the sex-ratio see the note on p. 312, above.

[82] Cf. the dissertation of Stark on the Geschlechtsverhältnis bei unehelichen Geburten und bei Totgeburten (Freiburg, 1877).

quantitative special classes of births that possess independent probabilities for the birth of a boy or a girl. Thus, according to Geissler's researches based on statistics of Saxony, more or less fruitful marriages are such classes; likewise, the influence of the residence (whether in city or country), the ages of the parents, the nourishment, etc., are perhaps not without influence. But all of these differences may persist in like proportions, so that a constant total probability results for the whole population.[83]

Aside from the sex-ratios of births, Lexis has shown that there is normal dispersion among the sex-ratios of deaths occurring in the lowest age classes, and to a degree also among those occurring in the highest age classes—but not among those in the middle classes. Consequently, for such age classes, in which physiological conditions are of primary influence on mortality, we may speak of a constant probability for the deaths of the two sexes. This evidently indicates a constant cause-complex, especially relating to the mortality of childhood, in consequence of which a definitely greater percentage of boys die than girls. We may then agree with Lexis' thesis "that the average resistance of boys to death is, upon organic grounds, less by a fixed

[83] Lexis in the article "Geschlechtsverhältnis der Geborenen" in the Handw. d. Staatsw., 2nd ed., p. 180; cf. also "Das Geschlechtsverhältnis der Geborenen und der Wahrscheinlichkeitsrechnung" in Lexis' Abhandlungen and the bibliography attached to the article cited above. Lexis has investigated (Theorie der Massenerscheinungen, pp. 74-78) the special relation existing among twin births and has found that the ratio of the number of twins in which both were boys to the number of twin (both) girls is the same as the ratio between single male and female births. The dispersion of these ratios for the older divisions of Prussia during the period 1862-1873 was very near normal, as was also the case for the percentage of mixed twin births. Aside from Lexis the sex-ratios among multiple births have been investigated by Westergaard, Geissler, Neefe, and Herrl.

amount than that of girls."[84] Further, the sex-ratios of the survivors of one and the same generation at the end of the first years of life can be cited[85] as likewise possessing normal dispersion and a typical value. It is to be noted that the relative numbers cited are all secondary (analytical, relative) numerical probabilities. Also the other cases of normal dispersion which can be established—even though merely for single countries and for definite periods of time—mostly relate to secondary probabilities. Thus, the percentage of females condemned for larceny (in relation to the total number condemned for that reason), which according to Westergaard[86] exhibited a normal dispersion in Denmark during the period 1867-1885, represents a secondary probability. Other illustrations are the ratio of female suicides to the total number, which Westergaard[87] found, in Denmark for the period 1861-1886 and in Belgium for the period 1865-1883, to coincide with the experience of games of chance, the relative frequency of suicide by hanging, and the relative frequency of suicide in the months October, November and December, for both of which Westergaard found normal stability for the Danish figures, 1861-1886.[88] In other countries the percentage of female

[84] Abhandlungen, p. 204. Cf. Geigel, Die Stabilität des Geschlechtsverhältnisses der Gestorbenen, 1880.

[85] W. Kammann, Das Geschlechtsverhältnis der Überlebenden in den Kinderjahren als selbständige massenphysiologische Konstante, Göttingen, 1900 (abstract in the Jahrb. f. Nat. und Stat., 3rd series, Vol. XIX (1900), p. 382 f.).

[86] Grundzüge der Theorie der Statistik, p. 50.

[87] See loc. cit., p. 44 f.

[88] Grundzüge, pp. 45 f., 47. v. Bortkiewicz finds ("Kritische Betrachtungen zur theoretischen Statistik," Conrad's Jahrb., 3rd series, Vol. VIII (1894), p. 672), in opposition to Westergaard, that the dispersion of the frequency of suicides by hanging (Westergaard's illustration) departs from theory by a not inconsiderable amount. Such a difference of opinion is possible because in judging the dispersion of a series from the standpoint of the theory of probability a certain subjective element always enters. v. Bort-

suicides and the classes of suicides according to the manner of death possess a smaller degree of stability. Lexis has shown, for instance, that in France the percentage of suicides by drowning exhibited a tendency to decrease for both sexes during the period 1835-1868, and that there is normal stability only for the female sex and that only for a part of this period selected for its small fluctuations. Lexis has further shown that the relative number of female suicides exhibited a tendency to decrease during that period.[89]

The most thoroughly investigated " primary " numerical probability in the demographic field is the probability of death. This probability for the years of childhood doubtless presents a decided supra-normal dispersion in its time fluctuations,[90] while from the age of ten years and up it presents a greater stability.[91] J. H. Peek[92] found dispersion differing but little from the normal (slightly supra-normal) for the male population of the Netherlands. A very good approximation to normal dispersion was found by Peek in the mortality ratios which have been deduced from the Dutch civil service statistics (1878-1894) for the construction of the first civil service mortality table for

kiewicz has, however, classified Westergaard's data of the frequency of suicide by hanging according to the sex of the suicide and has found that the dispersion for each sex is apparently normal, the dispersion for the female sex being unquestionably so.

[89] Cf. Theorie der Massenerscheinungen, pp. 83-87, and article "Gesetz" in Handw. d. Staatsw., 2nd ed., p. 239.

[90] Cf. Lewis, ibid. pp. 78-82.

[91] Tschuprow ("Die Aufgaben der Theorie der Statistik," Schmoller's Jahrbuch, 1905, p. 54 f.) ascribes the smaller degree of stability of the mortality of childhood as compared with that of later years of age to the greater influence that climatic conditions have upon the health of delicate children as compared with their influence upon adults and to the greater susceptibility of children to epidemics.

[92] "Das Problem des Risiko in der Lebensversicherung," Zeitschrift für Versicherungsrecht und -Wissenschaft, V (1899), pp. 169-197.

the Netherlands.[93] Normal dispersion and, therefore, the maximum stability of mortality ratios have been shown to exist among insured persons, by G. Bohlmann[94] upon the basis of the observations of the insurance societies of Gotha and Leipsic, and by E. Blaschke[95] upon the basis of the observations underlying the mortality table of the twenty British Societies from the year 1869. E. Blaschke has demonstrated the normal dispersion of the frequency of invalidity between the ages of 20 and 55 (upon the basis of Zimmerman's statistics of the invalidity of employees of German railways and Kaans's statistics of Austrian miners' friendly societies).[96]

Experience shows that other series of relative numbers based upon demographic or moral-statistical data do not, as a rule, possess normal dispersion, whether the items are for different years or for different geographical areas. Many such series have a definite evolutionary tendency, or exhibit either periodic fluctuations, or non-periodic variations caused from time to time by special economic or political events, all of these fluctuations being in evident contradiction to the theory of probability.[97] Even where

[93] Cf. Czuber, Wahrscheinlichkeitsrechnung, p. 333.

[94] Über angewandte Mathematik, Leipzig, 1900, p. 142.

[95] Cf. Die Anwendbarkeit der Wahrscheinlichkeitslehre im Versicherungswesen, Vienna, 1901, and Vorlesungen über mathematische Statistik, 1906, p. 144.

[96] Vorlesungen über mathematische Statistik, p. 144 f.

[97] For series or parts of series which exhibit a very slight tendency to increase or decrease we may overlook this tendency and proceed from the assumption of a constant probability. We may also attempt to explain an evolutionary series by assuming a probability which varies in a definite manner (increasing or decreasing), and we may measure the deviation for a given year from the hypothetical probability assumed for that year. Lexis, who has discussed this point (Theorie der Massenerscheinungen, p. 32; Abhandlungen, p. 192 f.), holds that the utility of such a proceeding is very doubtful. The accepted norm for the variation of the objective probability must always be arbitrary. If we represent this variation by means

there is no definite evolution and no periodic or other wave movement, the condition of the symmetric distribution of the items about the average is frequently not satisfied. Finally, in those cases where this condition is satisfied and the items are distributed about their average like accidental fluctuations the dispersion of the series is usually " supra-normal " and the stability " infra-normal," that is, the deviations from the average are greater than those called for by the theory of probability, and the probable or the mean error computed according to the " physical " method is often several times greater than the corresponding error computed by the " combinational " method, the precision computed by the former method is considerably less than it should be, theoretically. Series of infra-normal dispersion, that is, series in which the actual probable or mean deviation from the average is less and the actual precision is greater than that to be expected by the theory of probability, have up to this time not been discovered. We may assume that such an infra-normal dispersion or supra-normal stability could be exhibited only by mass-phenomena governed by some exterior restricting force.[98] However, it is to be remembered in this connection that the theory of probability sets a very strict standard where great numbers of observations are given. Series which, according to the theory of probability, possess a decidedly supra-normal dispersion, may appear to be extraordinarily stable to the non-mathematical statistician, who lacks a standard depending strictly upon the number of observations. Thus, the numerous series which, to certain writers, seemed to prove " regularity " and " law " among statistical phenomena,

of a curve or a broken line, then it is easy to draw the line or curve so that the deviations of the observed numbers from it are a minimum. It would be least objectionable to explain series having a uniform development as time progresses through assuming a probability likewise changing uniformly; such regular evolutionary series do not, however, actually occur.

[98] Cf. Czuber, Wahrscheinlichkeitsrechnung, p. 322.

have been most largely those series which, according to the theory of probability, do not possess normal dispersion. On the other hand, with a small number of observations the theory of probability allows for deviations of considerable size which are to be looked at as purely accidental, yet which may, perhaps, appear so considerable to the non-mathematical statistician as to require a special explanation.

Series of relative numbers in which the items are distributed symmetrically, yet with a supra-normal dispersion, are frequent.[99] Lexis,[100] who is supported by other mathematical statisticians,[101] explains such series by the assumption that the items are the result, not of a single constant probability as in the case of normal dispersion, but of a variable probability which is itself subject to accidental fluctuations. Series of supra-normal dispersion may, perhaps, be compared to the results of a game of chance, in which red and white balls are drawn from several urns, which do not contain red and white balls in the same, but in varying ratios, those ratios being themselves affected by accidental errors. There is a theoretical probability recognizable even in such series which, however, varies accidentally about a mean from one series of observations to another (which may relate to years, months, provinces, etc.).

The supra-normal dispersion of these series results because two different causes of errors are in operation,

[99] Thus Lehr has shown that in Germany during 1841-1885 the ratio of the number of still-births to the total number of births, as well as the ratio of the number of deaths to the total living population, exhibits deviations from the mean that are to be considered accidental disturbances, although the dispersion of the single items is considerably supra-normal. (Lexis, article "Gesetz" in Handw. d. Staatsw., Vol. V, p. 239.)

[100] Cf. Theorie der Massenerscheinungen, especially p. 26, and Abhandlungen, pp. 135 f., 176-184.

[101] Cf., for example, v. Bortkiewicz, Das Gesetz der kleinen Zahlen, § 14.

first, the "combinational" or "statistical" cause of errors, which is also connected with a constant probability and causes the deviations of the empirical values from the theoretical probability; second, the "physical" or "physiological" cause of errors, which originates from the fluctuations of the objective probability. These two causes of errors correspond to two fluctuation components, which Lexis calls respectively "normal-accidental" or "unessential," and the "physical" components, while von Bortkiewicz calls them the "normal error" and the "absolute excess of error." These two components cooperate in producing the total deviations, which are found by direct observation of the fluctuations.

If the dispersion or the stability of several series are to be compared, then, according to Lexis, only the physical fluctuation component is of essential importance. Series of normal dispersion (whose deviations from the mean depend exclusively upon the "normal-accidental" component) may indeed possess various degrees of variation, but this difference is the exclusive consequence of differences in the probability and in the number of observations, and the series compared coincide as to the constancy of their theoretical probability and the non-existence of a physical fluctuation component. Series of supra-normal dispersion, on the contrary, depend upon the simultaneous operation of the "normal-accidental" and a "physical" component of fluctuation. In order to compare two such series we must, according to Lexis, eliminate the normal-accidental components in both series from the values containing the total fluctuations and merely regard the physical fluctuation components which are independent of the number of observations, as it is only in these components that the degree of the variability of the respective probabilities of the series compared is expressed, that is, how great are the fluctuations to which these probabilities are subjected.

It has been established by Lexis[102] that approximately normal dispersion is exhibited especially by series of relative numbers, which are based on only a moderate basic number of observations and that the stability of statistical relative numbers, as a rule, varies inversely with this basic number. If the basic number is reduced by means of a time, space, or qualitative or quantitative subdivision of the statistical material it is not seldom that an unquestionably supra-normal dispersion is changed to an approximately normal one. This phenomenon may have two causes. It doubtless rests essentially upon the method of measuring dispersion, used by the mathematical statisticians, but, on the other hand, it may to a certain extent, as will be described later, depend upon the lack of interdependence of the elements belonging to a statistical mass, and the fact that these elements are influenced by "causes operating conjointly."

The influence of the method of measuring dispersion asserts itself in the following way: The measurement is effected—as explained above—by comparison of the values obtained for the probable or mean error (or the precision) computed, first, by the "physical" method and, second, by the "combinational" method. If the value found by the physical method is essentially greater than the other, then the dispersion is supra-normal in a degree determined by the difference between the values. In case of supra-normal dispersion the first value consists of two components, the normal-accidental and the physical, while the second value has but one fluctuation component, i. e., that corresponding to the normal-accidental component. The difference between the probable or mean errors (or the precision values) computed by the two methods depends, therefore, upon the ratio between the two fluctuation components named. Now the normal-accidental component, in conformity to the

[102] Cf. Abhandlungen, VIII, "Concerning the Theory of the Stability of Statistical Series," p. 187 f.

law of great numbers, varies inversely with the basic number of observations, while the physical component is, in general, independent of the basic number. Therefore, with relatively great basic numbers the physical fluctuation component prevails, while with relatively small basic numbers the normal-accidental component predominates. Consequently, if the accidental fluctuations, to which the objective probability underlying the series is subject, are not great, then with a moderate basic number the physical fluctuation component is almost wholly concealed and computation gives approximately normal dispersion, while the same statistical phenomenon founded upon greater basic numbers gives a decided supra-normal dispersion.

From the greater constancy of relative numbers with moderate sized basic numbers it does not follow, however,—as was emphasized especially by von Bortkiewicz,[103]—as a postulate of statistical research that we should be content with small basic numbers and base our conclusions upon the resulting statistical material. "It is, on the contrary, of greater statistical interest to establish the physical fluctuation-component, which is obscured by the use of small basic numbers. For its numerical magnitude is a measure, independent of the operation of "accidental causes," of the time changes which are experienced by the probability in question. At the same time it is to be observed whether the latter increases or decreases with time. But the amount and direction of the variation will be, in so far as it is a question of eliminating the accidental, so much more certainly ascertainable the more numerous are the observations upon which the relative numbers rest. Therefore, a limitation of the number of observations is not advisable. However, an investigation of statistical series with smaller basic numbers will recommend itself on account of those general theoretical interests which are concerned in demon-

[103] "The Theory of Population and Moral Statistics According to Lexis," Jahrb. f. Nat. u. Stat., Vol. XXVII (1904), p. 239.

strating cases of an approximately normal stability." The statistician is, therefore, correct in generally attempting to utilize great masses of observations, in which the accidental errors are eliminated, so that the non-accidental fluctuations or changes during the course of time may be established. If we desire, however, to investigate the relation of the law of chance to statistical data, that is, to determine whether the propositions and methods of the calculus of probability are applicable to statistics, then the conditions of the investigation must be so shaped that the effects of the factor, which is to be studied, are possible of evaluation.[104]

From this point of view von Bortkiewicz has continued the investigations of Lexis concerning the question of the greater constancy of relative numbers with moderate sized basic numbers. He obtained important conclusions and showed, in particular, that statistical series consisting of very small *absolute* numbers often exhibit stability almost completely satisfying the requirements of the theory of probability, and named this phenomenon "the law of small numbers" ("Gesetz der kleinen Zahlen"). As illustrating this law von Bortkiewicz used certain figures from suicide and accident statistics, figures for consecutive years representing very small "numbers of events," while, at the same time, each item originated from a great number of observations (that is to say, men under observation). The first illustration was based on the suicides of Prussian children under ten years during the period 1869-1893. The annual number of suicides of boys fluctuated between 0 and 6. For girls there were some years with no suicides; there was but one year in which more than one suicide was committed, and in that year there were but two suicides. There was coincidence between the actual dispersion found and the dispersion theoretically expected for boys and girls alone, as well as for both sexes combined. The second illus-

[104] Cf. v. Bortkiewicz, Gesetz der kleinen Zahlen, preface, p. v.

tration is the female suicides in the small German States of Schaumburg-Lippe, Waldeck, Lübeck, Reuss ä. L., Lippe, Schwartzburg-Rudolstadt, Mecklenburg-Strelitz and Schwartzburg-Sondershausen during the period 1881-1894. The annual number of suicides in these states fluctuated between 0 and 10 and the dispersion agreed with the theoretical distribution computed *a priori*. Similar results were given by the statistics of accidental deaths among the employees in eleven German corporations during the years 1886-1894 (the number fluctuated between 0 and 14, the numbers over 10 occurring each but once), and by the number killed in the German army by being kicked by a horse during the period 1875-1894.

This remarkable constancy of small numbers of events, as well as the constancy of relative numbers based upon small numbers of observations, doubtless essentially depends upon the method used to measure dispersion, which tends to conceal the physical fluctuation component of a variable probability if a small number of observations or events are given. Von Bortkiewicz has, however, an explanation of this constancy independent of the method of measuring dispersion in his theory of "causes operating conjointly."

In all investigations by means of the theory of probability we proceed from the assumption that the events are wholly independent of each other. This assumption is, however, a circumstance not always realized. To this fact is due the considerable difference that is so often observed in statistics between the actual and the expected (theoretical) groupings. But the same fact, with the help of the theory of "conjointly operating causes," also explains the greater constancy of the small event-numbers and that of relative numbers with small basic numbers. The interdependence of the single cases may be due to the peculiarities of the phenomenon in question. If we consider, for example, the accidents due to a boiler explosion or a mine disaster, the

individual deaths are not independent of each other, as in such catastrophes the lives of several persons are simultaneously destroyed. In such cases von Bortkiewicz speaks of an "acute" solidarity of the individual cases. A "chronic" solidarity of the individual cases occurs when their interdependence proceeds from influences such as climatic conditions, which affect all of the cases uniformly. Such solidarity exists, for example, among the deaths that occur in consequence of an extremely hot summer. Chronic solidarity of individual events may constitute the rule in statistics. If we combine a great number of observations where such solidarity exists the statistical masses will contain many interdependent units and the actual dispersion must exceed normal dispersion in a degree dependent upon the number of items bound together, while a closer approximation to normal dispersion will follow a limitation to smaller masses.[105]

Von Bortkiewicz has arrived at the conclusion that the statistical results satisfy the standard mathematical formulas better, if the field of observation is smaller or if the event in question is very unusual in a given society (such as suicide or accident).[106] The examples which von Bortkiewicz used to support his conclusion are not numerous, but he anticipates that others will be found and that they will help to confirm the scientific conviction "that mathematical probabilities or their functions underlie all numbers relating to population or moral-statistical phenomena." In this sense the law of small numbers appears to be a suitable "support for the explanation, in which statistical numbers are consequences of certain general conditions, to which accidental causes are added, and in this way new authority may be given to that theory of a specific statistical

[105] Cf. v. Bortkiewicz, Gesetz, Anlage 2, and Tschuprow, "Die Aufgaben der Theorie der Statistik," Schmoller's Jahrbuch, 1905, p. 57.

[106] Gesetz, p. 36.

regularity, which has been almost discredited by the blunders of Quetelet and his followers.''[107]

The remarkable permanence of certain small numbers of events was proven to be well-founded by Bowley, independently of von Bortkiewicz's *Das Gesetz der kleinen Zahlen*. Bowley cites as illustration the number of deaths in Great Britain from splenic fever.[108] These numbers for the period 1875-1894 were 5, 4, 10, 14, 12, 18, 9, 15, 8, 18, 11, 11, 11, 12, 7, 4, 3, 6, 7, 10. The average is 10; which is extremely minute in comparison with an annual average number of deaths amounting to 530,000, and which corresponds to an extraordinarily small probability. The fluctuations of the number of deaths from splenic fever are consistent with the theory of probability. There are other small numbers of events which, according to Bowley, often present extraordinary permanence, seldom increasing much and just as seldom entirely disappearing. ''Specialists in all professions, from the doctor who treats only one obscure disease of the ear, to the dealer in curiosities, make their livelihood dependent on this permanence of small numbers. The regular occurrence of accidents and improbable events in general furnishes other examples of the same sort.''[109]

[107] Ibid., preface, p. vi.

[108] Elements of Statistics, Pt. II, Section IV, "The Permanence of Certain Small Numbers," 2nd ed., p. 301 f.

[109] Ibid. p. 302.

APPENDICES

APPENDIX I

A. SERIES WHICH EXHIBIT A CHARACTERISTIC REGULAR DISTRIBUTION OF ITEMS IN OTHER WAYS THAN WITH RESPECT TO THE DISPERSION OF THE ITEMS ABOUT THEIR MEAN (SERIES OF CHARACTERISTIC CONFORMATION)

We have mentioned [1] several times the existence of series whose items have no regular dispersion about the mean but which have some other characteristic regularity and, represented graphically, give a definite characteristic curve. Such series cannot, evidently, be adequately represented by averages. The peculiar characteristic conformation of the series must be set forth. In order to accomplish this it is necessary to investigate the succession or order of the numbers and the relations between them, that is, to consider the entire conformation of the series, or the whole curve. Only in this way can the principle be found which lies at the basis of the relation of each item of the series to all the others.

It is our object to treat cursorily in the following of those series which we designate shortly as "series of characteristic conformation." A discussion of such series is not inappropriate in a work on averages, as the domain of averages is, to a certain extent, thereby limited negatively. The problem of representing and evaluating such series is frequently also positively connected with the problem of averages. Thus, the investigation of causes upon the basis of quantitative series of characteristic conformation may be considered an extension of the similar

[1] Compare pp. 268, 293, and 297 f.

investigation conducted by comparing averages and relative numbers. The discussion of the investigation of causes upon the basis of such series will be supplemented by a short explanation of the methods of the investigation of causes by means of a comparison of geographic and time series. Finally, the methods of measuring the correlation among several individual characters will be indicated, and thus the broad outlines of all the methods of statistical investigation of causes will be appropriately included in this book. The methods of the investigation of causes by means of the comparison of geographic and time series and the methods of measuring the correlation among several individual characters are connected with the problem of averages also by the fact that these methods make extensive use of averages for special auxiliary purposes.

Characteristic conformations are frequently found in time series of the second and third groups and also in quantitative series of the first and third groups.

Let us first consider *time series.* A characteristic conformation is shown by "evolutionary" series which present a definite tendency of development, and by "periodic" series which present definite regularly recurring time movements.

Evolutionary series have been established for definite periods of time in various statistical fields. Thus, Lehr found that in Germany during 1841-1885 the ratios of the number of births and the number of marriages to the population exhibited a tendency to increase uniformly with time, while the percentage of illegitimate births to the total number of births showed a tendency to decrease uniformly.[2] Wages of most modern countries show a steady advancement, but the birth rates of numerous countries have steadily decreased in recent years. An evolutionary series consisting of absolute numbers may, however, lose its evolutionary character if its members are changed to

[2] See Handw. d. Staatsw., article "Gesetz" by Lexis.

relative numbers; the quantity of a commodity imported may increase absolutely from year to year and yet the per capita amount imported may remain the same. On the other hand, an evolutionary series of relative numbers may change its character when the members become absolute numbers; thus, in a growing population with a constant annual number of births there is a continued decrease of the birth rate.

The significance of an evolutionary tendency naturally increases with its geographical extent and the length of time that it persists. An evolutionary tendency common to many countries is not seldom spoken of as a " statistical law." Thus, G. von Mayr speaks of a " law of the progressive undermining of the population in the central districts of great cities "[3] and designates the regularity with which urban elements increase and country elements decrease as a social law of development of modern times.[4]

Evolutionary series arouse interest in proportion to the steadiness with which the tendency of development is expressed, that is, in proportion to the infrequency of the fluctuations. The relatively great stability of certain evolutionary series has frequently led to their representation and characterization by mathematical formulas. As is known, it has been often asserted (first by Euler with reference to London) that population increases geometrically, that is, according to the compound interest formula.[5] Mathematical statisticians have not infrequently used other formulas to present population statistics or other data as functions of time.

But the agreement of the growth of population or other phenomena with a mathematical function is, obviously,

[3] Bevölkerungsstatistik, p. 63.

[4] Ibid. p. 61.

[5] Malthus draws a contrast between the "geometric" measure of the population and the alleged "arithmetic" increase of the means of subsistence.

merely accidental and is, as a rule, limited to short periods of time. Even if the tendency of development remains the same, frequent fluctuations of the rapidity of movement appear. Thus, it is never safe to assume that population will change in the same manner in the future as it did in the past. Therefore, the computations of the length of time during which a population could double itself, formerly common, have long been rightfully discarded. Nevertheless, it is necessary, in order to compute the number of population for an intra-censal year or for a post-censal year, to assume that the population changes regularly and to ascribe a definite law of development to it. It is chiefly for this purpose that mathematical statisticians have represented population as a function of time.[5a]

Periodic series as well as evolutionary series exhibit a characteristic conformation and can, therefore, not be adequately expressed by averages. The entire conformation must be investigated in order to reveal the periods of the series, the time when they appear, and their intensity.

The phenomena which exhibit certain regularly recurring periods are very numerous. The periods may be of various kinds, but seasonal fluctuations are most common. Most demographic and moral-statistical phenomena (mortality, births and marriages, crimes, suicide, etc.) all show the influence of the seasons more or less decidedly. This influence naturally varies with climatic conditions. A detailed investigation will also develop the facts that this seasonal influence does not affect all population groups uniformly, such as age classes, and that various causes of death, various diseases, various crimes, etc., possess seasonal curves peculiar to themselves. Likewise, various economic phenomena such as unemployment, consumption of various

[5a] An illustration of the use of the compound interest law to represent the "growth element" of certain financial statistics may be found in J. P. Norton's Statistical Studies in the New York Money Market.—TRANSLATOR.

articles, etc., are affected by seasonal fluctuations. In Austria the number of members of sick-benefit societies (depending on the intensity of employment) gives the same characteristic seasonal wave year in and year out, the trough of the wave coming in January and the crest in July or August.

Aside from seasonal periods there are also periods of longer duration in many fields of statistics. The regular succession of periods of economic expansion and depression is well known. This phenomenon is one which has interested many "crisis theorists," especially the statisticians Jevons and Juglar. Juglar has found periodic fluctuations among births and marriages of various countries that correspond to the economic fluctuations.[6] Other periodic movements are caused by the influence of certain days of the week and hours of the day upon certain phenomena.[7]

Some phenomena are simultaneously affected by several wave movements of various lengths; thus, many economic phenomena possess both seasonal fluctuations within the year and periods of expansion and depression covering several years. At the same time, many periodic phenomena are also subject to a definite evolutionary tendency. In addition, other disturbances are frequently caused by events which happen from time to time, such as wars, failure of crops, epidemics, etc. These various fluctuations may intermingle and hide each other. To establish the periodicity of a series and the length of its periods is, therefore, often a very difficult problem and its solution necessitates a thoroughgoing study of the series.

Series of quantitative individual observations and quanti-

[6] Compare "Y a-t-il des périodes pour les mariages et les naissances comme pour les crises commerciales?" Bulletin de l'Inst. int. de Stat., Vol. XIII, Pt. IV, p. 8 f.

[7] Compare, for example, Enrico Raseri, "Les naissances et les décès suivant les heures de la journée," Bulletin de l'Inst. int. de Stat., Vol. XI, Pt. I.

tative series of relative numbers or averages sometimes have a characteristic conformation. Of the first group of series those which present the structure of a population according to age frequently exhibit a characteristic conformation. G. von Mayr[8] distinguishes various characteristic forms of age structure when series of ages are represented graphically, such as the triangular structure of the population of Germany or the United States,[8a] the bell-like structure of the French, the onion-like structure of the population of great cities and industrial districts, and the spindle-formed structure of agricultural districts from which there is emigration to cities and industrial centers. Likewise, the age composition of those living according to the tables giving the number of survivors at various ages generally has a regular configuration. These numbers do not, indeed, result from direct observation but from computation; nevertheless, they are individual data and may be considered fictitious individual observations.

Other series of individual data to be considered are incomes and wages. Both incomes and wages of most countries give rise to series which, quite apart from the grouping of the items about their average, exhibit a characteristic regularity of conformation.

Quantitative series of characteristic conformation not infrequently arise from relative numbers and averages based upon various age classes. Thus, Quetelet found a regular curve for the rate of criminality according to age. The most important series of this kind is the probability of death according to age, which—at least between certain limits—gives a characteristic curve. Average wages for different age classes also frequently exhibit regular characteristic conformations.

[8] Bevölkerungsstatistik, p. 76 f.

[8a] Cf. United States Census Bulletin No. 13 for "A Discussion of Age Statistics," which gives the diagram of ages for the population of the United States.—TRANSLATOR.

The mathematical statisticians have often attempted to represent series of individual observations and quantitative series of the third group, which exhibit characteristic conformation, by means of *mathematical formulas and analytic functions.*[9] Thus, the number of survivors, the probability of death, or the frequency of sickness may be expressed as a mathematical function of age; the number of persons receiving a certain wage or income may be expressed as a function of the amount of wages or income received. Quetelet developed a formula in his *Physics of Society* which gave the frequency of crime as a function of age; he also represented the growth of men according to age by an equation of the third degree.

The best known formula for expressing mortality as a function of age is the one advanced by Gompertz in 1825 and later improved by Makeham.[10] Numerous other writers, such as Lambert, Babbage, Litrow, L. Moser, Edmonds, Lazarus, Opperman, Thiele, Wittstein, and others, have also developed formulas which express either the number of survivors or the probability of death as a function of age. Most of these writers have overrated the importance of the formulas that they developed. They thought to reveal a physiological law by means of their formulas. They supposed that the mortality curve possessed a definite general form and that, consequently, a mathematical law of mortality existed to which the mortality of every population fitted; the variations in mortality among different populations or groups of persons was

[9] "Every statistical table suggests the expression of the thing whose quantities it shows in one column, as a function of the thing whose quantities it shows in another." (Prof. Marshall in the article "On the Graphic Method of Statistics" in the Jubilee Volume of the Royal Statistical Society, 1885, p. 255.)

[10] It is $l_x = k \cdot s^x \cdot (g)^{c^x}$, where x is the age, l_x is the number living at this age, and k, s, g, and c are constants. See Harald Westergaard, Die Lehre von der Mortalität und Morbilität, 2nd ed., p. 201.

provided for in the various values for the constants which appeared in the standard mathematical formula.[11]

Modern statisticians, with few exceptions, have little faith in a general, uniform law of mortality. Mortality is evidently strongly influenced by the greatly divergent conditions, both natural and social, of various countries. It has also been demonstrated that mortality changes with time. It appears impossible, therefore, to find a formula for mortality which will be valid for the series of observations of all countries and all times.[12] However, mathematical formulas which satisfactorily describe definite mortality curves are of importance for such purposes as the computation of mortality, for reducing the labor of certain computations in life insurance,[13] for interpolation,[14] and for the adjustment of mortality tables.[15] The part which mathematical formulas play in the above cases is, however, quite different from the part that they would play as an expression of " natural law." In these cases the securing of a mathematical formula is not an end in itself but its application is merely a means toward attaining some concrete purpose.[16]

Vilfredo Pareto's mathematical theory of the distribution of incomes attracted much attention when it was published in his *Cours d'économie politique* (in 1896) and in other

[11] v. Bortkiewicz on "Gesetz der Sterblichkeit" in Handw. d. Staatsw.

[12] Emanuel Czuber, Wahrscheinlichkeitsrechnung, No. 197, "Mortality Formulæ."

[13] See v. Bortkiewicz, "Gesetz der Sterblichkeit."

[14] Compare, for example, Chap. X, "Interpolation," Section II, "Algebraic Treatment," in Bowley's Elements of Statistics.

[15] Thus, the first table of the civil service of Austria-Hungary was graduated by Makeham's formula.

[16] The objections that have been enumerated against a "law of mortality" also hold for B. Scratchley's mathematical formulation of a "law of sickness" in which the frequency of sickness is a function of age. (Compare with Westergaard, Mortalität und Morbilität, 2nd ed., pp. 89 and 201 f.)

writings. Pareto collected the income curves of several countries and times and found a mathematical formula which, by the insertion of appropriate constants, could be made to agree with all these curves.[17] "Here exists a law, a true law," said Foville of Pareto's formula,[17a] and, turning the law against socialistic tenets, he added: "Our presumptuous reformers will no more succeed in changing the natural curve of incomes than they can vary the parabolic paths of projectiles or the elliptic orbits of the planets." Foville is of the opinion that a change in the geometric shape of the income curve is not to be expected, but this view, he holds, does not eliminate the possibility of progress of the lower classes, since Pareto's formula contains variable parameters and the curve may become less steep in the course of time. Pierre des Essars, the French statistician, has tested Pareto's formula with the income statistics of Austria [18] and found the formula to hold also in this case.[18a] It is to be emphasized that Pareto's formula does not postulate the law of chance and does not depend upon the theory of error.

. Lucien March has brought the objection to Pareto's formula that it assumes the smallest incomes to be the most numerous.[19] This assumption he held to be untrue, as the income statistics of Saxony—which possess the peculiar feature of having no inferior limit—clearly show.

[17] This formula is $n = \frac{a}{x^{\alpha}}$; in which x signifies the income, n the number of incomes equal to or greater than x, and a and α are constants for any given series.

[17a] Économiste français, July 4, 1896.

[18] Compare Journal de la Société de Statistique de Paris, 1902, p. 222 f.

[18a] But see criticisms of Pareto's formula by W. M. Persons in "The Variability in the Distribution of Wealth and Income," Quarterly Journal of Economics, May, 1909, and by M. J. Séailles in La Répartition des Fortunes en France.—TRANSLATOR.

[19] See Journal de la Société de Statistique de Paris, 1902, p. 152 f.

March has himself offered a general formula for the distribution of a special category of incomes, namely wages.[20] He based the formula upon French, German, and American wage data and recommended (in the 1902 session of the Société de Statistique) that it be used to describe the distribution of all incomes. The curve which corresponds to March's formula shows the number of workmen in various wage classes. The minimum wage is not the most frequent and the distribution of wages about the modal or "normal" wage is unsymmetrical.

The views of mathematical statisticians are divided as to the value of these and other mathematical formulas for the representation of statistical series. We shall not enter into the various technical controversies, among which that concerning the degree of coincidence between the formulas and the statistical data is most important. It is, however, of general significance that most of the formulas constructed present only certain greater or lesser parts of the series in question; the formulas for mortality do not apply to the years of childhood;[21] Pareto's formula does not hold true for those receiving small incomes.

The principal question is independent of these questions of detail. Do such mathematical formulas express statistical laws and do they benefit the science? Lexis and von Bortkiewicz hold such formulas to be of little significance. Lexis pointed out that they merely present the exterior of a mass of items, and that only approximately. "Thus we may describe approximately the exterior surface of a sand heap by means of an empirical formula, but we would never consider this formula as the law which had controlled the

[20] See Journal de la Société de Statistique de Paris, 1898, pp. 193 f. and 241 f.

[21] As the practical application of this formula is chiefly in life insurance written principally for adults this objection is of little significance.

positions to be taken by the grains of sand." [22] Von Bortkiewicz says that the formula derived by Galton from Körösi's natality table fails to advance the knowledge of fecundity a single step; the friends of mathematical statistics have often pointed out the uselessness of similar formulas.[23] Lexis, von Bortkiewicz, and Edgeworth agree as to the decided inferiority of such empirical formulas to the formulas of the theory of probability, which latter offer a rational explanation of the grouping of the items about their average. However, we must remember that the series which statisticians have sought to represent by empirical formulas are the very ones that appear impossible to explain by the theory of error or of probability because of the irregular dispersion about the average. At the same time, such series exhibit a regular conformation which it is possible to characterize by a mathematical formula. Moreover, the formulas derived for quantitative series of characteristic conformation are to be considered mathematically precise laws of causation—as will be shown in the following chapter—and therefore they express relations of indubitable scientific importance.

The mathematical statisticians have, as we have mentioned, not only characterized series relating to single countries by mathematical formulas, but they have attempted to obtain formulas for certain phenomena that would be generally valid. For instance, they attempted to find general formulas for mortality according to age and for the distribution of incomes. They endeavored to formulate in mathematical terms the broad outlines exhibited by the respective phenomena in various countries and at various times and to express the peculiarities of the different countries and times by giving specific values to the constants of the mathematical functions. Such gen-

[22] Zur Theorie der Massenerscheinungen in der menschlichen Gesellschaft, p. 8.

[23] Jahrbuch für Nationalökonomie und Statistik, 1897, I, p. 127.

eral formulas are, however, very rare. If the broad outlines of certain phenomena are the same at various times and in different countries, such fact can of course also be established without the aid of higher mathematics. Moreover, it is doubtful whether the mathematical standard can be considered to settle the issue. Several series for which general mathematical formulas cannot be formed may, nevertheless, be regarded as of like conformation if compared according to the less rigorous standard of elementary mathematics. The series may then permit a conclusion as to uniformity of cause, whether these causes be biological or be rooted in the uniformity of the fundamental social institutions of various countries.

B. INVESTIGATION OF CAUSES UPON THE BASIS OF QUANTITATIVE SERIES OF CHARACTERISTIC CONFORMATION

Every series of regular conformation gives a picture of the association of two variables, but only in the case of quantitative series can an efficient causal relationship be deduced from examination of the conformation of the series.

Time series present the aspect of a phenomenon according to divisions of time. If the conformation is especially regular then it may be represented by a mathematical formula, in which the phenomenon presented by the series becomes a function of time. However, no deduction of a proper causal connection—for instance, the existence of a sociological relation—can be made.

In series of quantitative individual observations the two variables are, first, the element of observation in question and, second, the number of items belonging to each grade of the element measured. Such series show, for instance, the relation between the amount of income or wages and the number receiving specified amounts, or between age and the number in each age class. If the series in question presents a characteristic conformation, then the rela-

tion between the two variables concerned may be definitely formulated, perhaps by a mathematical formula in which the frequency of the items becomes a function of their magnitude. But the interrelation characterized in this way is merely a mathematical one; a proper causal connection does not exist between these two variables any more than it does with time series. Thus, the amount of the income or wages received does not *cause* any particular number of individuals to receive such income or wages, and age is not the *cause* of the unequal frequencies of the age classes. The conformation of such series is merely descriptive in its significance. A mathematical function can merely give the contour and the extent of the masses of items without discovering or defining a causal nexus.[24]

Quantitative series of the third group are quite different from time series and series of individual observations in this respect. The two variables concerned are, first, that quantitative characteristic which is used to differentiate the constituent masses, and second, the numbers characterizing those constituent masses, both of which are values which give properties of definite statistical masses and, therefore, can stand in a direct causal relationship to each other. If a quantitative series—or a considerable part of one—exhibits a regular conformation it signifies that some relation of dependence exists between the two variables. Such regular conformations result when the numbers characterizing the constituent masses consistently increase or decrease as the mark of differentiation increases. Examples are, respectively: the increase of the probability of death with age, and the decrease of fecundity with better economic position. The hypothesis that such regularity originates accidentally can usually be ruled out. The regu-

[24] In the theory of error the number of the deviations of items from the mean is expressed as a function of the size of the deviations. Here again a purely mathematical relationship is created which does not reveal the cause of the deviations.

larity of the conformation indicates a definite causal law. In the examples cited above the conformation shows a definite influence of age upon the probability of death, and of economic condition upon fecundity.

Series of such regularity that the variables change in a definite ratio (such as that defined by the equation $y = ax + b$) are extremely rare. It is only by way of exception that cause and effect can be brought into direct relationship and there are usually causes, other than the one referred to in the series, which disturb the regularity. But if it is possible to etablish that one variable varies directly or inversely with the other, this is of much significance and justifies a conclusion. Sometimes it is possible to divide a series into several parts, in each of which a peculiar relationship between the variables may exist. The direct parallelisms may become inverse. Thus, in childhood the probability of death decreases as age increases, while in later years it increases with age, and wages increase with age up to the time of maximum efficiency and then decrease.[25-25a]

[25] Austrian data provide an interesting illustration of the connection beween age and wages. The information collected concerning the conditions of miners in the Ostrau-Karwin coal district shows that miners between the ages of 36 and 40 years receive the highest incomes. Previous to that age group incomes increase with age (although not uniformly) and decrease thereafter. The decrease from the highest paid age class is, however, very gradual and less in amount than the increase from the youngest age classes. (See Arbeiterverhältnisse im Ostrau-Karwiner Steinkohlenreviere, published by k. k. Arbeitsstatistisches Amt im Handelsministerium, Pt. I, p. 56, and graphic representation of Table VIII, following p. 38.)

A similar result is given by the wage statistics of industrial workers of northern Bohemia. "The wages of men increase until the age class 31-35 years is reached, remain stationary in the succeeding age classes to 45 years, and then decrease with increasing ages, in spite of the advancement of a number of the male workmen into the better paid positions of foreman, inspector, and master

The relations between the pairs of variables may be very diverse, and their formulation may introduce various mathematical functions. Mathematical formulas which represent the conformation of quantitative series are, by their nature, mathematically precise statements of cause. Consequently, such formulas possess more scientific value than those which indicate the conformation of time series or series of individual observations without revealing any causal nexus. To be sure, the exact mathematical formulation of complicated functional relationships possesses no great value for sociology, especially if no further explanation can be given of the type of connection thus revealed.

The investigation of causes on the basis of quantitative series of characteristic conformation is subject to the same

workmen. Accordingly, the age of maximum earnings of workmen is from 31 to 45 years." "Among the women workers the greatest efficiency and, therefore, the highest wage comes somewhat earlier than for men, i. e., between the ages 25 and 35 years, and for time wages, peculiarly enough, in the still earlier age class, 21 to 25 years. The wages of women do not rise and fall with age as closely as do those of men." (Nordbohmische Arbeiterstatistik, Ergebnisse der von der Reichenberger Handels- und Gewerbekammer am 1. Dezbr. 1888 durchgeführten Erhebung, Erläuterungen zu den Tabellen, p. xxxvii.)

F. W. Lawrence has made some interesting investigations as to the connection between wages and the size of the city in which the workmen reside. He has shown from wage data of the building and printing trades and iron manufacture that, in general, wages increase with the size of the city because, indeed, the "demands for social life" increase with the size of the city where the workman lives. (See Local Variations in Wages, London, 1899.)

[25a] Volumes I to V, inclusive, of the United States Bureau of Labor Report on Condition of Women and Child Wage-Earners in the United States give wage data tabulated according to age of the worker. For instance, in the incandescent electric lamp establishments, employing 2,430 women, there is a rapid rise of wages from age 16 to age 20, a gradual rise from age 20 to age 24, and, finally, a fall from age 24 to the group, 45 and over, when the 17-year-old wage level is nearly reached.—TRANSLATOR.

restrictions and involves the same presuppositions as the investigation of causes by comparison of two averages or relative numbers.[26] A primary consideration is that the parallel or opposite changes of the two variables merely signify a connection between them, without specifying which is the cause and which is the effect. This question may be a difficult one and must be settled by some non-statistical method. Frequently, the causal connection between the two phenomena is not one-sided but mutual. Under certain conditions the two phenomena may not be in direct causal relationship but the correspondence which they exhibit may be due to their dependence upon a common, and, perhaps, unknown cause.

The method of investigation of causes upon the basis of quantitative series, like the method of comparison of two averages or relative numbers, postulates *ceteris paribus.* If the constituent parts which form the series are distinguished also by some characteristic other than the criterion used to differentiate them, then it is impossible to specify the cause of the differences of the magnitudes characterizing such constituents. For example, if the mortality of children was shown to increase with the number of children per family, but if the larger families were also the poorer ones, then the greater mortality might be due to greater poverty as well as to larger families.

With reference to the *ceteris paribus* hypothesis the method of investigation of causes on the basis of quantitative series is superior to that of the comparison of two averages or relative numbers. If but two values are compared (for instance, mortality in two occupations), then; in addition to the criterion used in distinguishing the two masses, there may be an indefinite number of other differences between the two masses which might cause the difference noted between the values compared. However, if a

[26] Compare with the chapter on "Investigation of Causes by Comparison of Averages and Relative Numbers," p. 110.

quantitative series is of regular conformation, then only those additional causes which act in a similar regular manner can complicate the problem. Consequently, it appears that a large number of causes must be ruled out and a conclusion in regard to causation is made considerably easier.

For example, suppose we are investigating the mortality of two groups of the population which are distinguished both by difference of economic condition and of occupation. Suppose the mortality rates are different. It is impossible to ascribe a precise cause for this difference, as it may be either economic condition or occupation. However, if we investigate the mortality of a series of groups defined by economic condition and obtain a regular curve, then we may be able to conclude that economic condition is the cause of the differences in mortality, even if each group according to economic condition contains different occupations. That the regularly graded mortality is due to the different occupations appears extremely improbable. The regular shape of the curve could be due to occupation only in case the occupations belonging to the different economic groups are, by chance, arranged according to the degree of mortality. This is extremely improbable; if necessary, light may be obtained by special investigations of another kind.

Only when there is question of a second cause exhibiting the same quantitative gradations as those resulting from the criterion applied is there difficulty in arguing the presence of a causal nexus from the existence of a regular curve. Thus, in the illustration cited above of the mortality of children varying with the size of the family, and the size of the family varying with economic condition, the greater mortality rates may be due either to the larger families or to greater poverty. Frequently, however, there is no question of a second cause and the existence of a regular quantitative series indicates that the regularity is

due to the criterion at the basis of the grouping and not to some unknown additional cause. Other peculiarities of the constituent masses may, indeed, be present. But these express themselves, as a rule, only through disturbances of the regular form of the curve.

The method of investigation of causes upon the basis of quantitative series of regular conformation is, in a way, an extension of the method of comparison of single averages or relative numbers. Conclusions which are drawn from the comparison of two values may be controlled, further developed, and corrected by the formation of a whole series of values. Thus, the difference in mortality between city and country is well known. But is it possible to maintain that mortality changes constantly with the size of the population group? Ballod has shown that in Prussia mortality according to age classes is most unfavorable in the middle-sized cities, then follow the large cities, the small towns, and, finally, the country.[27] The size of the population group appears to be of influence, but there is no consistent parallelism. Another illustration follows.

Some statisticians have maintained that the difference of ages of parents has an influence upon the sex of the offspring. Körösi has tested this question by observations in Budapest, with the result that in those cases where the fathers were decidedly older or younger than the mothers the percentage of boys born was considerably greater, but the sex-ratio of the children did not vary uniformly with the difference of ages of the parents.[28]

[27] Bulletin de l'Institut intern. de Statistique, Vol. XIV, No. 1, p. 135. See also Carl Ballod, Die mittlere Lebensdauer in Stadt und Land. (Staats- und sozialwissenschaftliche Forschungen, edited by Prof. Schmoller, Vol. XVI, Pt. V, 1894.)

[28] Neue Beiträge zur Sexualproportion der Geburten. (Bulletin de l'Institut intern. de Statistique, Vol. XIV, No. 4, p. 14.)

C. INVESTIGATION OF CAUSES THROUGH COMPARISON OF GEOGRAPHIC AND TIME SERIES

Investigation of causes upon the basis of quantitative series presupposes the formation of graded constituent parts defined by a quantitative criterion. Many times, however, there are insurmountable obstacles to the prosecution of a concrete investigation by the method described in the preceding section. In such cases we may have recourse to an " indirect " method. Instead of directly applying the quantitative criterion in question to form the constituent parts of a series, we may make use of available geographic or time masses, if these masses vary at the same time with respect to the quantitative element in which we are interested. For example, suppose that we are investigating the connection between economic well-being and mortality. If there are statistical difficulties which prevent a classification of both the population and deaths according to economic well-being, then we may examine a series based upon geographic divisions, which divisions contain populations of various degrees of well-being, with reference to their mortality. Or, we may compare various periods of time, during which economic conditions are varied, with reference to mortality. Which " indirect " method we choose is a question of expediency. Many times both methods are equally practicable. Geographic and time series must also always be utilized if the various grades of the element in which we are interested are not to be found at the same time or in the same country.

However, it may happen that geographic or time series do not give the solution of quite the same problem to which the comparison of quantitative constituents of the same totality relates. In such cases the indirect method possesses independent significance, but of course it cannot be used as a substitute for the direct method. Thus the influence of economic well-being will be found to be essentially different in case we study the various degrees of well-

being which exist permanently for various social classes as compared with those which result from quickly alternating periods of industrial depression and expansion affecting the whole population.

If we depend upon a *geographic series* to establish a certain causal relationship we have to arrange the geographic divisions of the series according to the magnitude of both quantitative elements between which the dependence is supposed to exist. Thus, we could easily arrange the series according to mortality and according to economic well-being. If the two characteristics are really causally connected, then the geographic divisions in question will, on the whole, appear either in the same order or in opposite order in both arrangements according as the relation is direct or inverse.

Such comparisons of geographic series are quite frequent and numerous causal relationships are thus ascertained. The comparison of series fulfils its purpose as completely, however, when an expected relationship is contradicted as when a causal nexus is demonstrated. Comparisons of geographic series have been made to ascertain the connection between economic well-being and mortality, well-being and number of children, birth rates and mortality of children, size of farms and density of population, mortality and density of population, etc. It has been repeatedly established in England that the most densely populated registration districts also have the highest mortality rate. The earlier English statisticians saw in this parallelism of increasing density of population and rate of mortality the expression of a general statistical law. However, this parallelism is not exhibited by the more recent statistics of other countries.[29]

Geographic series based upon density of population and upon birth rates have also been compared. The question whether the marriage rate and the frequency of illegitimate

[29] Compare G. v. Mayr, Bevölkerungsstatistik, p. 222 f.

births are connected has been advanced by J. Bertillon. *A priori* one might expect that where the marriage rate is higher fewer illegitimate children would be born. A comparison of geographic series has not confirmed this expectation.[30] J. Bertillon has likewise investigated the question of the relation between the frequency of illegitimate births and age of marriage. The comparison of such series in fact indicated a connection between a higher age at marriage and a greater frequency of illegitimate births, and conversely.

Numerous other illustrations of the application of these methods might be given. Among these there are some cases which have been interpreted variously, thus leading to controversies. Thus, Achille Guillard's "Loi du Rapport inverse," according to which growth of population stands in inverse relation to density of population, caused much dispute. In his *Eléments de Statistique Humaine, ou Démographie Comparée,* which appeared in 1855, Guillard arranged 114 countries and provinces, first according to their respective densities of population and, second, according to the rate of growth of their population, and he found that, on the whole, the geographic divisions appeared in opposite order. The violence with which Wappäus, for example, has opposed this law is interesting.[31] Only one of Wappäus's numerous arguments in opposition will be cited. He contended that it was fallacious for Guillard to base his conclusions upon averages for great geographic areas consisting of heterogeneous parts, such as the density of population of Russia. Such averages, which may differ widely from the individual conditions which they represent, are quite frequently used in comparisons of geographic series (for instance, when nations are used as units), and consequently the conclusions are often question-

[30] Compare J. Bertillon, Cours élémentaire de Statistique administrative, p. 480.

[31] Allgemeine Bevölkerungsstatistik, Pt. I, p. 144 f.

able. This objection to the method of comparison of geographic series can, however, be overcome by using " natural districts," when possible, instead of political divisions. Detailed investigations may enable us to form homogeneous geographic complexes, of a greater or smaller size, which correspond to the various gradations of a definite quantitative characteristic (for example, natural areas for various degrees of mortality). These statistical districts which correspond to the various degrees of a definite phenomenon may then be examined with reference to any other phenomenon in which we are interested to see if the two phenomena are related. If the second phenomenon exhibits the same—or the inverse—arrangement of areas, when the items of the two series are placed according to magnitude, it means that the two phenomena are causally connected. G. von Mayr was able to form well defined districts in Bavaria distinguished by various degrees of child mortality. He then ascertained the density of population for these areas, and the birth rate and the frequency of illegitimate and still-births as well, and in this way investigated the relationship between these phenomena.[32] In a similar manner we could form " natural districts " according to density of population or size of farms and ascertain the amount of emigration from these divisions to cities or to foreign countries, or the degrees of any other phenomenon which might be dependent upon density of population or the distribution of land ownership.

Time series are compared just as frequently as are geographic series in order to determine whether their conformations are similar and whether a causal connection can, therefore, be assumed. The coincidence of two time series may be either in their evolutionary tendency or in their concomitant or synchronous oscillations. Whenever a causal nexus is assumed, whether the two series show parallel or opposite variation, we speak of the items as corre-

[32] Compare G. v. Mayr, Theoretische Statistik, p. 88.

lated quantities. The comparison of time series is of interest even though there is no question of a causal connection between the two phenomena. We might investigate the question whether there is a progress in the economic well-being of a country corresponding to the per capita increase of taxes (per capita imports plus exports, or per capita consumption being used as indices of well-being). Or, we might inquire if any selected phenomenon (such as foreign trade) exhibits the same evolutionary tendency in various countries, etc.

The comparison of the course of two or more time series is greatly facilitated by graphic representation. The synchronous oscillations of "historical curves" are much easier to locate than similar variations in the numerical data. Graphic representation is, therefore, strongly advised particularly for "experimental" investigation of causes. If various types of phenomena (such as foreign trade, marriage rate, unemployment, etc.) are represented in the same diagram, then the impression of the diagram and its conclusiveness, of course, depend upon the scales chosen for the several curves and the relation of the scales to each other.[32a]

A certain space of time must often elapse between the operation of cause and effect. In such a case the movements of the two curves are not synchronous but are separated by that part of the curve corresponding to the elapsed time. Thus, R. H. Hooker[33] has found that in England the parallelism of marriage rates and foreign trade during the period 1861-1895 is not greatest for simultaneous fluctuations but for marriage rates and imports (or total foreign trade) which precede the marriage rates by a third of a

[32a] Cf. Bowley, Elements of Statistics, Chap. VII, "The Graphic Method," for a discussion of the comparison of curves and the choice of scales.—TRANSLATOR.

[33] "Correlation of the Marriage-Rate with Trade." (Jour. of the Royal Stat. Soc., 1901, p. 487 f.)

year, or exports which precede marriage rate by about half a year. Hooker found that a year and a quarter intervened between total bank clearings and marriage rates.

The earlier statistical text-books usually gave the frequently observed correspondence between the number of births and marriages and the price of grain as an illustration of correlated time series. The price of grain was taken as an index of economic conditions. However, it is no longer an accurate index of conditions and, consequently, the parallelism that formerly held true has not existed for some decades. Phenomena which have been more recently used as indices of the economic conditions and the movements of which have been compared with each other and with births, deaths, marriages, etc., are the state of the labor market, the per capita amount of foreign trade, the amount of savings, the consumption of certain articles, the percentage of poor receiving government aid, the per capita amount of bank clearings (the last appears in the official annual Report of the English Registrar General), etc.[33a]

The fluctuations of criminality and the changes in the price of grain have likewise been compared, and thus the well-known parallelism between the price of grain and larceny, and the inverse relation between the price of grain and bodily assault, have been established. But these relations no longer exist, probably for the reason that the price of grain no longer serves as a barometer of the economic condition of the great masses of the population. The same statement holds true of the coincidence between grain prices and the number of German emigrants, which items moved together previous to 1870. A great number of influences,

[33a] A number of series of economic statistics showing synchronous fluctuations have been charted by Mr. W. H. Beveridge in his Unemployment. The chart is given the significant title "The Pulsebeat of the Nation."—TRANSLATOR.

which have nothing to do with the price of grain, now affect the number of emigrants.[34]

Of more recent date are Juglar's investigations concerning the occurrence of economic crises at the times when bank loans were highest and bank reserves lowest. In order to indicate what a great variety of series may be compared it is sufficient to mention the attempt to explain the often observed simultaneous appearance of sun-spots and economic crises; the demographic congress of 1887 considered this question as well as that of the relation between sun-spots and mortality. An English author has even discussed the connection between English death rates and the orbital motions of the planet Jupiter.[35]

Just as it is the tendency of other methods of modern statistics to depend more and more upon minute investigations, so in the comparison of time series more attention is now paid to the details of the phenomena. It is especially practicable to take into consideration the differences between various social classes. If we are investigating a causal connection it is evident that, where possible, only those masses should be compared which, on the one hand, express cause and, on the other hand, express effect. To include in the comparison masses which do not participate in the causal connection must evidently spoil the picture and make the demonstration more difficult. G. von Mayr has emphasized this idea in connection with the special problem of the relation between criminality and the price of grain. He makes the point that, evidently, " millionaires are not immediately driven to larceny by an increase in the price of grain "; " it is, therefore, not sufficient to compare total criminality with the price of grain; if we wish to get valuable results we must pay especial attention to

[34] Compare G. v. Mayr, Bevölkerungsstatistik, p. 347.

[35] B. G. Jenkins, "On a Probable Connection between the Yearly Death-rate and the Position of the Planet Jupiter in His Orbit." (Jour. of the Roy. Stat. Soc., 1879, p. 330.)

the criminality of various social classes.''[36] It is just as evident that the size of the crop and the price of grain must exert different influences upon the producing and the consuming classes, whether we are examining the movement of the population, marriages, or crimes. Nevertheless, this difference was almost neglected by the earlier writers. By distinguishing between producers and consumers B. Pokrovsky has obtained valuable new results in his study of the influence of crops and the price of wheat upon the movement of population in Russia,[37] as has also Dr. J. Buzek, who investigated the influence of crops and prices of grain upon the movement of the population of Galicia (a province of Austria) during the period 1878-1898.[38]

If the similarity of the fluctuations of two series is established, then we may proceed to investigate the ratio between the fluctuations. For instance, is the change in the marriage rate always a certain percentage of the change in the amount of exports? As a rule we must be satisfied if we find that greater fluctuations in one series are accompanied by greater fluctuations in the other series, as exact proportions seldom exist.[39]

If an exact expression of the degree of parallelism between two series is desired, then we may look for a numerical measure which shall vary with the degree of similarity between corresponding fluctuations of all the items of the two series. Thus, pairs of series may be graded according to the degree of parallelism. March and the English statisticians have paid especial attention to the derivation of

[36] "Über die statistischen Gesetze." (Bull. de l'Inst. intern. de Stat., Vol. IX, No. 2, p. 309.)

[37] Given at the St. Petersburg session of the International Statistical Institute in 1897. See Bull. de l'Inst. intern. de Stat., Vol. XI, No. 1, Pt. II, p. 176.

[38] Statistische Monatsschrift (Vienna), 1901, pp. 167-216.

[39] Bowley has described a special graphic method of testing the proportionality of the fluctuations of two series. (Elements of Statistics, 2nd ed., p. 177.)

such a numerical measure. The former has constructed a "*coefficient de dépendance*" upon elementary mathematical formulas. By means of this coefficient, the correspondence of the fluctuations of two time series is summarized and an index is obtained which varies with the degree of coincidence of the series compared.[40] The English mathematical statisticians have developed a "coefficient of correlation," based upon the calculus of probability, which varies with the degree of correlation between the two series compared. Numerous ingenious mathematical studies of the theory of correlation have been made.[41-41a] Whether any two phenomena are dependent or not is judged by the magnitude of the coefficient of correlation computed for them.[41b]

[40] See "Comparison numérique de courbes statistiques." (Journ. de la Soc. de Stat. de Paris, 1905, pp. 255 f. and 306 f.)

[41] Compare Pt. II, Section VI, "The Theory of Correlation" in Bowley's Elements of Statistics. The most noteworthy writers on the subject are Edgeworth, Pearson, Galton, Yule, Hooker, and Sheppard.

[41a] For a résumé (mathematical) of the work in this field, together with some new applications, see "The Correlation of Economic Statistics" by W. M. Persons in the Quar. Publics. of the Am. Stat. Assoc. for December, 1910. For an extensive mathematical treatment of frequency-curves, dispersion and correlation, see Yule's An Introduction to the Theory of Statistics (1911). For a non-mathematical explanation of the meaning of the standard deviation, coefficient of correlation, etc., see W. P. and E. M. Elderton's Primer of Statistics (1909).—TRANSLATOR.

[41b] The coefficient of correlation was first derived by A. Bravais in 1846, who, however, did not use a separate symbol to stand for the function, nor did he apply it to statistics. Galton, Pearson, Edgeworth, and Yule have applied the function to statistics and developed the theoretical side.

The coefficient of correlation was derived by assuming that a large number of independent causes operate upon each of two series, producing normal distributions in both cases. Upon the assumption that the set of causes operating upon the first series is *not independent* of the set of causes operating upon the second series a

Certain authors have used the theory of correlation also to express the relationship between three variables. Social phenomena are, as a rule, dependent upon very many causes. Instead of relating a phenomenon to a single cause—as is commonly done—we may seek to determine its dependence upon two or more causes. Thus, Hooker and Yule have estimated the influence of price and of the quantity of

function is found of the form $\frac{\Sigma(xy)}{n\sigma_1\sigma_2}$. This is the coefficient of correlation (r). The notation is as follows:

X = any measurement in the first series.

Y = any measurement in the second series.

x = deviation of any item of the first series from the arithmetic mean of the series.

y = deviation of any item of the second series from the arithmetic mean of the series.

σ_1 = standard deviation of the X series.

σ_2 = standard deviation of the Y series.

n = number of items in each series.

It has been demonstrated that r cannot be greater than + 1 nor less than — 1. Positive values of r mean that large items of the X series occur simultaneously with (are paired with) large items of the Y series; negative values of r mean that large values of the X series occur with small values of the Y series. A value of r approximating O means that there is no correlation between the two series. There can be perfect positive correlation (r = + 1) or perfect negative correlation (r = — 1) only in case the items of the two series are connected by a function of the first degree (graphically, a straight line).

Karl Pearson has found the *probable error* of r to be 0.67 $\frac{1-r^2}{\sqrt{n}}$. A. L. Bowley has laid down the following rule for judging the meaning of r: "When r is not greater than its probable error we have no evidence that there is any correlation, for the observed phenomena might easily arise from totally unconnected causes; but when r is greater than, say, 6 times its probable error, we may be practically certain that the phenomena are not independent of each other, for the chance that the observed results would be obtained from unconnected causes is practically zero." (Elements, 2nd ed., p. 320.)—TRANSLATOR.

wheat produced upon the wheat exports from India.[42] The mathematical treatment, of course, becomes considerably complicated when more than two variables are used.

The investigation of causes by means of the comparison of geographic and time series is subject to similar limitations as the investigation of causes by means of quantitative series of characteristic conformation, and by the comparison of two averages or relative numbers.[43] Which of two apparently related phenomena is cause and which is effect cannot be determined by the statistical comparison. The two phenomena may coincide because they have a common cause or causes,[44] or each may react upon the other. Mutual influences are known to be especially common among social and economic phenomena. Thus—to give but two such cases that may be investigated by means of time series —there is a reciprocal action between prices and consumption, and between freight rates and tonnage.

Further, it is to be borne in mind that conclusions in regard to causation based upon geographic and time series are always hypothetical, that is, they always rest upon the assumption *ceteris paribus*. However, the investigations of causation by the method of comparison of series and upon the basis of regular quantitative series are in a better position as regards this assumption than is the method of the comparison of averages or relative numbers. If a number of geographical divisions are arranged according to the intensity of each of two criteria and if the items of

[42] "Note on Estimating the Relative Influence of Two Variables upon a Third." (Journ. of the Roy. Stat. Soc., 1906, p. 197.)

[43] See above, pp. 110 f. and 353 f.

[44] The parallelism between divorce and suicide established by J. Bertillon upon the basis of geographical comparison should be mentioned in this connection. The countries with a high suicide rate show, in general, a relatively high divorce rate. This parallelism does not rest upon a direct causal connection, but may, however, result from common causes. Dipsomania, for example, leads to the increase both of suicide and of divorce.

both series appear in the same order, then, as a rule, we may assume the existence of a causal connection between them, since it is extremely improbable that the selfsame arrangement occurs in both by chance. Also, in the case of the comparison of time series which vary together and exhibit corresponding fluctuations it is difficult to find any valid explanation other than that of a direct or indirect causal connection between two phenomena. That two independent phenomena should give rise to similar curves with synchronous fluctuations during any considerable time would be extraordinarily improbable.

D. CORRELATION BETWEEN INDIVIDUAL CHARACTERS

The investigation of the association of two individual characters (elements of observation, individual measurements) with reference to the correlation existing between them is essentially allied to the investigation of causation by comparison of two time series.[45] Two individual characters are said to be correlated when the variations of one character are, on the whole, matched by corresponding variations of the other character. In such a case a direct causal connection may be present, so that the changes in one variable may be the cause of the changes in the other; but there may be a common cause producing the fluctuations in both characters. It is possible to investigate the correlation between two characters only in case they belong to the same individuals observed, or if the measurements of one phenomenon may be paired with measurements of another phenomenon. The second case is similar to the case of two time series in which every item of the first series is paired with an item of the second series relating to the same period of time (year, month, etc.).

[45] Such data of observation are presented in a "correlation table," which is a double frequency table, i. e., each element appears simultaneously in two classes of measurement.

Measurements of two or more characters of the same individuals are frequently utilized for the measurement of correlation in biology and anthropology. For example, if the lengths of two different parts of the body are measured for a group of individuals we may determine whether, as a rule, those individuals which possess a larger character of the first kind also possess a larger character of the second kind. If this is the case the two characters measured are positively correlated. As a matter of fact, research has shown that biological and anthropological characters are usually correlated to a sufficient degree, so that changes in one character are accompanied by definite positive or negative change in other characters.[46] The greatest degree of correlation is exhibited by the right and left parts of animals. Thus, Galton has applied the method of measuring correlation to the numbers of Müllerian glands on the right and left sides of swine. The number of these glands varies widely with the individual. In case of absolute symmetry there would be the same number of glands on both right and left sides and the correlation would be perfect. As a matter of fact, although the correlation is not perfect it is very great.

A table giving the ages of bride and groom at marriage offers an illustration of combined observations not belonging to the same individuals but still suitable for the application of the test for correlation. The two variables are the ages at marriage of the two sexes. The problem may be stated as follows: does the age of the groom at marriage vary in a definite way with the age of the bride at marriage?[47]

Much light has been thrown upon the problems of hered-

[46] Compare Georg Duncker, Die Methode der Variationsstatistik (1899), II, "Correlation," III, "Some Problems of Statistical Method."

[47] Compare G. Udny Yule, on the "Theory of Correlation." (Journ. of the Roy. Stat. Soc., Vol. LX (1897), p. 813.)

ity and selection by recent applications of the theory of correlation. Pearson and Galton have been pioneers in this field, where observations belonging to different individuals are paired. Galton, for instance, has compared the stature of parents and children and found that the average stature of the sons born of fathers of a given stature is nearer to the average stature of the population than is the stature of the fathers. This phenomenon was called "regression." The coefficients of correlation and regression afford a method of measuring the force of heredity and the effects of natural selection.[48]

The same mathematical methods which are being applied to the measurement of correlation between the items of two time series can also be applied to the association of two individual characters.[49-50] But the problems which are dealt with in this way are, by no means, peculiar to mathematical statistics. They likewise confront the statistician who merely uses elementary mathematics. Of course he will be obliged to make approximate judgments and will not be able to get such precise results as can be obtained by the refined mathematical methods.

[48] Compare especially Galton's Natural Inheritance and Pearson's "Mathematical Contributions to the Theory of Evolution," III, "Regression, Heredity, and Panmixia." (Philosophical Transactions of the Royal Society of London, A, 1896, Vol. CLXXXVII.)

[49] The most important works on the correlation of individual characters have been contributed by Galton, Edgeworth, Pearson, Weldon, and Yule. Compare also Georg Duncker, Die Methode der Variationsstatistik (1889), II, "Correlation."

[50] See also the references given on p. 367. Karl Pearson's Grammar of Science contains a very clear explanation of the correlation of individual measurements.—TRANSLATOR.

APPENDIX II

QUETELET'S "AVERAGE MAN"

Quetelet declares in his *Physics of Society* that he has undertaken the task of determining the man "who is to society what the center of gravity is to bodies." This is the "average man," a fictitious being "in whom all processes correspond to the average results obtained for society," "the mean about which the elements of society oscillate."[1] Quetelet's average man possesses in an average measure the physical characteristics and the mental attributes both of the people and the period which he represents. All of his characteristics and attributes are "in a proper equilibrium, in a perfect harmony, equally removed from exaggerations and defects of every sort, so that he must be regarded (for the period in question) as the type of all that is beautiful and good."[2]

Quetelet's average man has, as is well known, given rise to much critical discussion. First, the opinion of Quetelet that the average man represented the type of the beautiful and the good was generally rejected. The average man

[1] Über den Menschen und die Entwicklung seiner Fähigkeiten, oder Versuch einer Physik der Gesellschaft, German edition by Riecke, Stuttgart, 1838, p. 15.

[2] Ibid. p. 575. On p. 570 Quetelet expressed himself as follows: "If the average man could be completely determined, then, as I have already remarked, he could be considered as a type of the beautiful; and all considerable deviations from his proportions and from his qualities and capacities are to be ranked as deformities and disease; whatever feature is not only dissimilar to the proportions and forms corresponding in him, but is still more extreme than the cases observed, would be ranked as a monstrosity."

can by no means be regarded as physically beautiful. "The average color of the eyes would not meet the demands of beauty, and the average profile would surely be far removed from the ideal; besides in the majority of men the physical characters fall short of the standard of beauty (round shoulders, flat chests, warts, and excrescences)."[3-3a]

Just as little can the average man be regarded as the ideal moral type. He possesses all the good moral attributes only in the average measure, but besides these also all the bad moral attributes in a certainly not inconsiderable measure. Quetelet thinks, to be sure, that "an attribute of man becomes a virtue when it is a golden mean, which is equally far removed from all extravagances and which is between the limits marking the beginnings of vice." But this view does not take into consideration the numerous reprehensible human attributes. Quetelet's praise of moral mediocrity would only be justified if it were established that an excess of the average of one good quality were insepar-

[3] Westergaard, Die Grundzüge der Theorie der Statistik, p. 276. See also J. Bertillon, Cours élémentaire de Statistique administrative, p. 117, and A. de Foville, "Homo medius" (a paper read before the XIth Session of the Intern. Statis. Institute in 1907, and printed in the Bull. de l'Inst. int. de Stat., Vol. XVII, p. 46). In opposition to these authors G. Viola has recently taken Quetelet's position, i. e., that the average man corresponds in his physical proportions to the ideal of beauty. ("La teoria dell' 'uomo medio' e la legge dell variazioni individuali," Rivista italiana di Sociologia, 1906.)

[3a] See Hankins' very interesting study of Quetelet as Statistician (published as a Columbia University Study in History, Economics, and Political Science). In this study Quetelet's position is set forth and criticised. Joseph Jacobs has published two papers in which he attempts to build up an average Englishman and an average American. (See "The Middle American" in the American Magazine for March, 1907, and "The Mean Englishman" in the Fortnightly Review, Vol. LXXII, p. 53.) Mr. Jacobs assigns his average man a birthplace and early history, a household budget, an occupation, definite personal qualities, religion, political affiliation, etc. —TRANSLATOR.

ably united with a defect in some other good quality or with the excess of a bad quality.

The question now arises, what value may be given the average man for special statistical purposes apart from his alleged significance as a type of the beautiful and the good? Quetelet's average man is in a way the bearer of all the averages which may be determined statistically for a definite population and a definite time. By means of him comparisons of different countries and times may be established; and, furthermore, by the comparison of individual cases with the average man a standard for the judgment of these cases may be obtained. Thus, Quetelet thinks that in medicine "the consideration of the average man is important to this extent that it is almost impossible to judge the condition of an individual without comparing it with that of a fictitious being who is regarded as normal and who is in reality nothing but the average man whom we have in mind. A physician is called in to see a patient and finds upon examination that the pulse is too quick or the respiration too hurried, etc. It is evident that when such a judgment is made, we recognize that the observed phenomena diverge not only from those of the average man or man in the normal state, but that they pass the danger limits. Every physician in making such a judgment relies upon the data in the possession of science or upon his own experience, that is to say upon a computation of the kind which we wish to see carried out on a larger scale and with greater exactness."

Quetelet has correctly apprehended the principal purposes of averages: to make comparisons possible and to afford a standard for the judgment of individual cases. But in all statistical comparisons we have to do with some single observation element, and the comparison is accomplished by relating two individual averages or relative numbers—for example, the average stature of the inhabitants of two countries, or the rate of mortality in two countries. Simi-

larly, in the judgment of an individual case we have to do with a single definite observation element—for instance, the determination as to whether the stature or the length of life of a particular individual is above or below the average, and if so, how much. Hence, while the comparison of individual averages and their application as a standard for the judgment of items have the greatest methodological significance, on the other hand the summing up of all conceivable averages into one fictitious average man has no real statistical value whatsoever. If we had to compare the average men of two countries, we should have to resolve them into the individual averages from which they were constructed, and then compare these averages with each other. In the same way, in order to compare a single individual with the average man, we should have to consider one by one the various individual characters (stature, length of life, etc.) and to estimate the value of each by comparing it with the average value ascribed to the average man.

Let us inquire further whether the construction of Quetelet's average man, apart from any practical utility, is possible at all. We must distinguish here between individual characters such as stature, length of life, etc., and statistical magnitudes which, by their nature, express not the qualities of individual men but the frequency of definite events (births, deaths, crimes, etc.) for definite groups of men.

It is not inconceivable that a man may possess a number of individual characters in an average measure—for instance, average height, weight, muscular strength, income, length of life, etc. But how would it be in regard to those characters which not all individuals possess? For such characters there are no general averages which refer to the total population and hence they cannot be ascribed to the average man. For example, an average wage cannot be ascribed to the average man, since many men receive no wages at all. The same reasoning would hold true of the average age

at marriage. Accordingly, the average man cannot be characterized in regard to many points, and in these respects he cannot be used either for comparisons or as a standard for the judgment of individual cases.

Thus, the construction of an average man with individual characters meets with great difficulties, especially since many attributes, such as mental capacities, cannot be measured at all, though the question is, in this case, debatable. It is, however, quite impossible to equip this average man with those averages which express the mean frequency (or probability) of certain events (frequency of births, crimes, suicides, probability of death, etc.). These are values which are produced by interrelating statistical masses and have a meaning only in connection with the masses from which the events in question proceed. If, for instance, the frequency of crime amounts on the average to 1.2°/₀₀ for the entire population, this means that of 10,000 inhabitants 12 commit a crime in a year. But the "average man" either commits a crime or does not. In the first case his average would be 1,000°/₀₀, in the second 0°/₀₀.[4]

Mention should also be made of the question, broached by some authors, as to whether a man possessing all the observation elements in an average measure may really exist. Quetelet has himself expressed the opinion that the perfect type of the average man is hardly within the bounds of possibility; that in general it is only possible to attain the type in single, more or less numerous, respects.[5] It is, in

[4] The "abstract man" of Lexis is to be distinguished from the "average man." (See "Übersicht der demographischen Elemente," Bull. de l'Inst. de Stat., Vol. VI, p. 40, and Abhandlungen, p. 60.) The former is not characterized by definite attributes, but he merely functions as the bearer of various demographic probabilities. It is, according to Lexis, the final goal of demography to compass the life history of man, considered in the abstract. v. Bortkiewicz designates Lexis' "abstract man" to be a revised and improved edition of the "average man." (Conrad's Jahrb., 3rd series, Vol. XXVII, p. 245.)

[5] Über den Menschen, etc., p. 576.

truth, highly improbable that an individual should correspond to the average measure in various respects at once. Whether the averages of two observation elements will coincide more or less often in the same individuals depends essentially upon the extent of the correlation between those elements.

As previously mentioned,[6] two individual characters are in correlation if their sizes are related. In case of perfect correlation between two characters the averages must appear simultaneously; if two characters are measured in the same individuals (for instance, the length of two different parts of the body), then, in case of perfect correlation, those individuals who have one character of average size must also have the other of average size. If there is only a partial correlation between the two characters the averages need not appear simultaneously in all cases, but there is an increased probability of such appearance and the degree of this probability depends upon the degree of correlation.

Now, in fact, in biology and anthropology there is generally at least a partial correlation between various characters and, therefore, there is an increased probability for the simultaneous appearance of the averages of these characteristics. On the other hand, there is a lack of correlation between demographic and economic elements of observation as well as between these and anthropological elements.[7] Of course no one will claim that length of life increases in a definite ratio with bodily size or with age at marriage. There is, therefore, no increased probability for the simultaneous appearance of the various averages of

[6] Compare p. 370 f.

[7] The only known illustration of an undoubted correlation offered by demography is the combined ages of those marrying. But in this case we are not concerned with two characters of the same individual, but by the paired observations of different individuals, namely, one individual of each sex contracting marriage.

these elements. People of average height will not be found relatively more often among those of average length of life than among those of any other age. It follows from this that people who unite even a few averages for different observation elements (not exclusively in anthropology) are quite exceptional.

This fact must unquestionably diminish the value of the "average man." While single averages may often possess a typical character, that is, may be looked upon as normal values, from which the concrete single cases only diverge because of accidentally disturbing causes, the average man who combines all the averages can by no means be looked upon as a typical normal man from whom all other men only diverge accidentally. He is an abstraction without any foundation in fact and has no independent methodological value.

The "average man" has evidently sprung into being rather too hastily from Quetelet's generalizations upon the results obtained by him in anthropology. If all observation elements, like measurements of height, showed a symmetrical distribution about a typical mean and if, besides, there were a considerable degree of correlation between the various observation elements, then the average man would be indeed a good representative of the prevailing characters. But since these two presuppositions are not fulfilled, the theory of the average man is hardly more than a historical reminiscence, of interest only in connection with the personality of its author. In modern statistics, which emphasizes specialization and detail work, the average man cannot have any additional significance. Average values for whole nations are seldom useful for scientific purposes; the great object is to obtain values for smaller masses, definitely characterized and as homogeneous as possible, in order to form a basis for comparisons or for other scientific investigations.

APPENDIX III

REFERENCES

The following are the more important systematic and general methodological works on statistics, inclusive of general works on population statistics:

BAILEY, W. B., Modern Social Conditions, New York, 1906.

BENNINI, R., Principii di Statistica methodologica, Turin, 1906.

BENNINI, R., Principii di Demografia, Manuali Berbera, 1901.

BERTILLON, JAQUES, Cours élémentaire de Statistique administrative, Paris, 1896.

BLOCK, M., Traité théorique et pratique de Statistique, 2nd Ed., 1886.

BLOCK-SCHEEL, Handbuch der Statistik, Leipsic, 1879.

BOWLEY, ARTHUR L., Elements of Statistics, 3rd Ed., 1908.

BOWLEY, ARTHUR L., An Elementary Manual of Statistics, London, 1910.

CHEYSSON, E., Les méthodes de la Statistique. (Reprint from the Revue du Service de l'intendance militaire), Paris, 1890.

COLAJANNI, NAPOLEONE, Statistica e demografia:
  I. Statistica teorica, 1904.
  II. Demografia, 1904.

CONRAD, J., Grundriss zum Studium der politischen Ökonomie, Part IV, Statistik, 2nd Ed. Part I: Die Geschichte und Theorie der Statistik; Die Bevölkerungsstatistik. Part II: Statistik der wirtschaftlichen Kultur; 1st Division: Berufsstatistik, Agrarstatistik, Forst- und Montanstatistik. 2nd Division: Gewerbstatistik.

DAVENPORT, C. B., Statistical Methods, New York, 1904.

DUFAU, Traité de statistique, 1840.

ELDERTON, W. PALIN, Frequency-Curves and Correlation, London, 1906.

ELDERTON, W. P. and E. M., Primer of Statistics, 1909.

FALLATI, Einleitung in die Wissenschaft der Statistik, Tübingen, 1843.

FARR, W., Vital Statistics, London, 1855.
FAURE, FERNAND, Éléments de Statistique, Résumé du cours fait à la Faculté de Droit de Paris, 1904-1905, Paris, 1906.
FERROGLIO, GAETANO, Elementi di statistica teorica, 2nd Ed., Turin, 1891.
FIRCKS, A. v., Bevölkerungslehre und Bevölkerungspolitik, Leipsic, 1898.
FLECHEY, E., Notions générales de Statistique, Paris, 1872.
FLUX, A. W., Article on "Statistics" in Palgrave's Dictionary of Political Economy.
FRANCHI, L., Appunti di statistica, Modena, 1898.
GABAGLIO, ANTONIO, Storia e teoria generale della statistica, Milan, 1880.
GARNIER, JOSEPH, Éléments de Statistique, contained in Notes et petits Traités, Paris, 1865.
GUILLARD, ACHILLE, Éléments de Statistique humaine ou démographie comparée, Paris, 1855.
JACQUART, CAMILLE, Statistique et science sociale, Brussels, 1907.
JAHNSON, Theory of Statistics, 2nd Ed., St. Petersburg, 1887 (Russian).
JULIN, ARMAND, Précis du Cours de Statistique, Paris, 1910.
HAUSHOFER, MAX, Lehre- und Handbuch der Statistik, 3rd Ed., Vienna, 1904.
HAUSHOFER, MAX, "Bevölkerungslehre" in Natur und Geisteswelt, Vol. 50, 1904.
HAASE, HERMANN, "Die Statistik als Hilfsmittel der Sozialwissenschaften," in Kultur und Fortschritt, Nos. 341, 342, 1911.
JONAK, E., Theorie der Statistik in Grundzügen, Vienna, 1856.
KING, W. I., The Elements of Statistical Method, New York, 1912.
LEVASSEUR, E., La population française, Paris, 1889.
LEXIS, W., Einleitung in die Theorie der Bevölkerungsstatistik, Strassburg, 1875.
LEXIS, W., Article on "Statistik (1 Allgemeines)" in the Handwörterbuch der Staatswissenschaften.
LEXIS, W., Zur Theorie der Massenerscheinungen in der menschlichen Gesellschaft, Freiburg i. B., 1877.
LEXIS, W., Abhandlungen zur Theorie der Bevölkerungs- und Moralstatistik, Jena, 1903.
LIESSE, ANDRÉ, La Statistique, ses difficultés, ses procédés, ses résultats, Paris, 1905.
LONGSTAFF, G. B., Studies in Statistics, London, 1891.
MAJORANA, CALATABIANO G., Teoria della statistica, Rome, 1889.
MAJORANA, CALATABIANO G., La statistica teorica ed applicata, Manuali Barbera, 1889.

MATHESON, R. E., The Mechanism of Statistics, Dublin, 1889.

MAYO-SMITH, RICHMOND, Article on "Statistical Method" in Palgrave's Dictionary of Political Economy.

MAYR, GEORG v., Statistik und Gesellschaftslehre, Vol. I, Theoretische Statistik, 1895; Vol. II, Bevölkerungsstatistik, 1897; Vol. III, Sozialstatistik, 1909, 1910 (not yet complete).

MEITZEN, AUGUST, History, Theory and Technique of Statistics, translated from the German by R. P. Falkner, and published by The Macmillan Co., for the Am. Econ. Ass'n.

MESSEDALGIA, ANGELO, La Statistica e i suoi metodi, Rome, 1876.

MINGUEZ, y Vicenta Manuel Don, Tratado de Estadistica, Cordova, 1899.

MISCHLER, ERNST, Handbuch der Verwaltungsstatistik, Vol. I, Stuttgart, 1892.

MOREAU, DE JONNÉS, Éléments de Statistique, 1847.

NEWSHOLME, A., The Elements of Vital Statistics, London, 1892.

NIGGL, A., Grundzüge der Statistik, Leipsic, 1902.

NINA, L., Principî fondamentali di statistica, Turin, 1907.

OETTINGEN, ALEXANDER v., Die Moralstatistik in ihrer Bedeutung für eine christliche Socialethik, 3rd Ed., Erlangen, 1882.

PIDGIN, Practical Statistics, Boston, 1888.

PRINZING, FRIEDRICH, Handbuch der medizinischen Statistik, Jena, 1906.

QUETELET, A., Über den Menschen und die Entwicklung seiner Fähigkeiten, oder Versuch einer Physik der Gesellschaft, German translation by V. A. Riecke, Stuttgart, 1838.

QUETELET, A., Letters on the Theory of Probability, translated from the French by O. G. Downes, London, 1849.

RAMERI, LUIGI, Elementi di statistica, Turin, 1896.

RÜMELIN, G. (and v. Scheel), "Die Bevölkerungslehre" (Schönberg's Handbuch der politischen Oekonomie, 4th Ed., Tübingen, 1896, Vol. I, Pt. I) and "Die Statistik"; also in Schönberg's Handbuch.

SALVA, D. M., Tratado elemental de Estadistica, Madrid, 1882.

SCHNAPPER-ARNDT, G., Socialstatistik, Vorlesungen über Bevölkerungslehre, Wirtschafts- und Moralstatistik, Leipsic, 1908.

STEIN, LORENZ v., System der Staatswissenschaft, two volumes, Stuttgart, 1856, Vol. I, System der Statistik, Populationstatistik und der Volkswirtschaftslehre.

SÜSSMILCH, J. P., Die göttliche Ordnung in dem Veränderungen des menschlichen Geschlechts, etc., Berlin, 1741.

TAMMEO, G., La Statistica, Turin, 1896.

TUCKSCHERER, T. B. H., Statistische Onderzoek bei Nyghen van Detmar, Rotterdam, 1909.

TURQUAN, VICTOR, Manuel de statistique pratique, Paris and Nancy, 1891.

VERRYN-STUART, C. A., Inleidung tot de beolfening der Statistik, Part I: De statistische Methode, Haarlem, 1910.

VIRGLI, FILIPO, Statistica, 4th Ed., Milan, 1907.

WAGNER, ADOLF, Article on "Statistik" in the tenth volume of the Bluntschli-Brater Staatswörterbuch.

WAPPÄUS, J. E., Allgemeine Bevölkerungsstatistik, Göttingen, 1859-61.

WAPPÄUS, J. E., Einleitung in das Studium der Statistik.

WESTERGAARD, HARALD, Die Grundzüge der Theorie der Statistik, 1890.

WESTERGAARD, HARALD, Die Lehre von der Mortalität und Morbilität, anthropolog-statistische Untersuchungen, 2nd Ed., Jena, 1901.

WIRMINGHAUS, A., Article on "Statistik" in Elster's Wörterbuch der Volkswirtschaft.

WORMS, RENÉ, La Statistique, Revue intern. de Sociologie, 1904, No. 7, or in Philosophie des Sciences sociales, Pt. II (Méthode), Paris, 1904.

YULE, G. UDNY, An Introduction to the Theory of Statistics, London, 1911.

ZALESKI, W., Teorya statystyki w zarysie, Warsaw, 1884.

ZANZUCCHI, F., Lezioni di statistica, Parma, 1898.

ZANZUCCHI, F., Appunti di statistica, Parma, 1900.

ZUCCAGNI, ORLANDINI ATTILIO, Elementi di Statistica, Firenze, 1869.

INDEX